UNIVERSITY OF CALIFORNIA

Los Angeles

Big Frog in a Small Pond:

Undermatching Status, College major, and Their Influence on Early Career Earnings

A dissertation submitted in partial satisfaction of

the requirements for the degree Doctor of Philosophy

in Education

by

Shuai Li

2020

ABSTRACT OF THE DISSERTATION

Big Frog in a Small Pond:

Undermatching Status, College major, and Their Influence on Early Career Earnings

by

Shuai Li

University of California, Los Angeles, 2020

Professor Mark Kevin Eagan, Chair

Traditionally, colleges and universities have been expected to promote social mobility (Haveman & Smeeding, 2006). It has been widely recognized that higher education is one of the best investments an individual can make. Greater focus now has been placed upon where individuals actually went to college, instead of simply whether one went to college or not. The relationship between college selectivity and earnings has been demonstrated by the fact that higher selectivity is generally associated with higher earnings (Hoekstra, 2009; Beyond, Brewer, Eide , & Ehrenberg,1999). In addition to the fact that earnings are associated with college selectivity, the major field of study students choose is also influential. As a factor that has long been recognized, college major exerts great influence on college graduates' labor market outcomes (Rumberger &Thomas, 1993; Thomas 2003). However, there lacks empirical studies that explores the influence of postsecondary undermatching on students' labor market outcomes, and especially the different influence of undermatching in different academic field (STEM and non-STEM). Therefore it is essential to understand the role college major plays when studying the effect of undermatching on students' labor market outcomes.

Therefore, this study examines *who*, *how* and *what* of the relationship between undermatching and choosing a STEM major. The design of this study was guided by two sets of conceptual framework, including the college decision framework adapted from Perna (2006) and Iloh (2018), and human capital theory (Becker, 1975; Mincer, 1957). Guided by these two frameworks, the study conducted several multilevel analyses (HGLM, HLM), utilizing data from three sources, including the Educational Longitudinal Study of 2002 (ELS 2002), American Community Survey (ACS 2005), and Integrated Postsecondary Data System (IPEDS).

Findings reveal that the influence of undermatching on students' early career earnings does differ by academic major: for students choosing a non-STEM major, attending a less selective institution probably is not a good idea; however for students that chose a STEM major, sometimes being a "big frog in small pond" might actually be beneficial economically. Still, considering the prevalent undermatching rate and low STEM rate, especially among underrepresented minority and low-income students, K-12 education and higher education stakeholders should make concerted effort to ensure that students attend higher education institutions that best fit them, and that higher education institutions provide sufficient resources for them to succeed. The study then concludes with recommendations for K-12 and higher education policy and practice.

The dissertation of Shuai Li is approved.

Cecilia Rios-Aguilar

Michael Seltzer

Craig Enders

Mark Kevin Eagan, Committee Chair

University of California, Los Angeles

2020

Table of Contents

List of Figures and Tables

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support and help from many wonderful people. Firstly, I wish to express my sincere appreciation to my committee: Professor Kevin Eagan, Professor Michael Seltzer, Professor Cecilia Rios-Aguilar, and Professor Craig Enders. I appreciate all the thoughtful feedback and suggestions you provided during the proposal and dissertation process. Kevin, you have been a great advisor, your suggestions were consistently on-point, challenged me to be better, and my doctoral journey would not have been completed without your guidance, support and encouragement along the way. Mike, you are always so nice to students, and I am very grateful to your insights that helped me resolve the methodological issue. Cecilia, thank you for taking the initiative to reach out to me right before my proposal and making my transitioning process as smooth as possible, and I also appreciate your suggestions on the conceptual side of the dissertation. Craig, thank you for the opportunity to learn from you, working with you made me learn so much and it has been an honor. All in all, having you all on my committee is truly amazing!

I also would like to pay special gratitude and respects to professor Robert Rhoads. Rob, I was really lucky to have you as my advisor. You were the most dedicated mentor and you impacted me in the most profound ways, and I will forever miss you.

Many thanks to the HEOC family- I have learnt so much from you all, and I will cherish our time together.

Lastly, to my parents and husband, thank you for being there for me no matter what. Your love is invaluable and always gives me strength.

VITA

2012	B.A., Sociology Sun Yat-sen University, Guangzhou, China
2013	M.A., Education University of California, Los Angeles, California
2015	Research Assistant Graduate Division Diversity, Inclusion and Admissions office University of California, Los Angeles, California
2015-2017	Research Assistant Higher Education Research Institute University of California, Los Angeles, California
2017-2018	Research & Special Project Intern Dashew Center for International Students & Scholars University of California, Los Angeles, California
2017-2019	Teaching Assistant Department of Psychology University of California, Los Angeles, California

SELECTED PUBLICATIONS AND PRESENTATIONS

Rhoads, R. A., Li, S., & Ilano, L. (2014). The Global Quest to Build World-Class Universities: Toward a Social Justice Agenda. *New Directions for Higher Education, 2014*(168), 27-39.

Li, S. (2014, April). *Factors Influencing College Students' political aspiration: An Analysis by Racial Group*. Paper presented at the annual meeting of the American Educational Research Association, Philadelphia, PA.

Ilano, L., & Li, S. (2014, October). *The Global Quest to Build World-Class Universities: Toward a Social Justice Agenda*. Poster presented at the annual meeting of the Association of Study in Higher Education, Washington D.C.

Luo, L, Pojar, A, & Li S. (2020). *Dashew Center Programs: Utilization for 2016-2017. Institutional Report,* Dashew Center for International Students and Scholars (DCISS).

Chapter 1: Introduction

The concept of college undermatching has caught the attention of higher education researchers and stakeholders. During the former Obama administration, a White House report advocating for increasing college access for low-income students recognized this problem: "…too few low-income students apply to and attend colleges and universities that are the best fit for them, resulting in a high level of academic undermatching – that is, many low-income students choose a college that does not match their academic ability" (White House Summit on College Education, 2014, p. 4). To be more specific, postsecondary undermatching refers to the situation in which students choose to enroll in a higher education institution that is less selective than the ones they could enroll, given their academic credentials, such as GPA and SAT scores. But more research is needed to better understand the complexities of undermatching, including labor market outcomes. Accordingly, this dissertation seeks to understand what factors predict undermatching, how undermatching influences individuals' earnings, and how college major moderates such influence.

The undermatching phenomenon is not rare. A large number of high-achieving, low-income high school graduates do not attend a selective college or university, and some do not even apply to one (Hoxby & Avery, 2013; Smith, Pender, & Howell, 2013; Belasco & Trivette, 2015). It is estimated that among all high school graduates, the overall undermatching rate is at least 30% (Belasco & Trivette, 2015); in North Carolina, scholars have estimated the undermatching rate to be approximately 60%, while in Chicago Public School district the figure is 40% (Rodrick et al., 2009; Bowen, Chingos, & McPherson, 2009).

Traditionally, colleges and universities have been expected to promote social mobility (Haveman & Smeeding, 2006). It has been widely recognized that higher education is one of the best investments an individual can make. Sources have shown that attaining a higher education degree leads to greater success in the labor market (NCES, 2012). With the current situation that 66% of high school graduates enroll in postsecondary institutions (Bureau of Labor Statistics, 2014), the hierarchical structure within the higher education system increasingly exerts a crucial influence on the distribution of life chances. Greater focus now has been placed upon where individuals actually went to college, instead of simply whether one went to college or not. The relationship between college selectivity and earnings has been demonstrated by the fact that attending an institution of higher selectivity is generally associated with higher earnings (Hoekstra, 2009; Beyond, Brewer, Eide , & Ehrenberg,1999). Hence, it can be argued that when highly qualified students from low-income backgrounds undermatch, they ultimately may limit the potential for higher education to be a force in support of social mobility.

As discussed by Goldthorpe and Jackson (2008), a merit-based higher education system should offset the influence of social class in determining economic success, serve as a filter that prevents parents' economic position from directly being passed on to their children, and thus promote social mobility. However, in contradiction of promoting social equality and equalizing life chances among all students regardless of socioeconomic background, the selection processes within higher education seems quite removed from the social mobility related goals. There are signs that postsecondary institutions are increasingly comprised of youth coming from advantaged backgrounds; moreover, this pervasive phenomenon appears more frequent in selective institutions (Haveman & Smeeding, 2006; Dowds & Bensimon, 2005; Bastedo & Jaquette, 2011; Posselt, Jaquette, Bielby, & Bastedo, 2012).

Moreover, if students choose to attend a college the selectivity that matches their measured academic ability, there is a higher chance that they will complete a degree, and this is true for students of all academic ability levels (Light & Strayer, 2000). Peers, student body makeup, and campus culture typically influence students' learning habits, and degree aspirations (Smith et. Al., 2013). Also, less selective institutions typically tend to have smaller budgets and a lower expenditure per student (Hoxby, 2009), which could result in less academic and non-academic support, and ultimately lower graduation rates. Students who attended less selective institutions feel less academically challenged, perceive less benefit from attending college, and are less satisfied with their experience than students who attend institutions matching their prior educational achievement (Fosnacht, 2015). In short, undermatching limited students' capability development, and restricted their upward mobility.

In addition to the positive correlation between institutional selectivity and post-college earnings, students' major field of study students also tends to have a positive association with earnings after college. As a factor that has long been recognized, college major exerts great influence on college graduates' labor market outcomes (Rumberger &Thomas, 1993; Thomas 2003). According to Pascarella and Terenizini (2005), major can explain 25 to 35 percent of the earnings differences of college graduates. Specifically, Science, Technology, Engineering, and Math (STEM) college graduates are reported to have higher income than students majoring in other disciplines. Also, in a 2012 report, the annual income for graduates in engineering and computers and mathematics was $55,000 and $46,000 respectively, while the figure for humanities and liberal arts was $31,000 (Carnevale, Cheah, & Strohl, 2012). Therefore it is essential to understand the role college major plays when studying the effect of undermatching on students' labor market outcomes.

As a research topic that initially attracted great attention from the educational research community approximately a decade ago, relatively few articles exist that empirically explore the influence of postsecondary undermatching on students' outcomes. Generally speaking, previous studies argue that students attending a more selective college have a higher probability to graduate (Horn, 2006; Smith, Pender, Howell, & Hurwitz, 2012) and earn higher salaries (Loury & Garman, 1995; Thomas, 2003; Dale & Krueger, 2011). One study examined the mechanism through which undermatching shapes the first-year college experience among high-achieving students by using two national surveys about college experience; this study adopted the method of propensity score matching, and found out that undermatched students experienced lower levels of academic challenge, perceived lower satisfaction towards campus life, and experienced fewer gains (Fosnacht, 2015). Another study (Goodman, Hurwitz, & Smith, 2014) conducted an analysis of SAT takers in the 2004-08 graduating high school cohorts in Georgia, using a regression discontinuity design, and showed that attending a less selective institution decreases the probability of bachelor's degree completion. Still, the long-term consequences and implications of undermatching, especially relative to labor market outcomes, are not adequately explored.

In order to fill the gap in understanding how undermatching influences students' labor market outcomes, this study aims to answer the following research questions:

1. How do students' background characteristics influence their chances of undermatching?
2. How do students' background characteristics, especially undermatching status, influence their chances of choosing STEM majors?
3. To what extent does undermatching influence students' labor market outcomes (i.e. annual earnings from employment)?

3.a. To what extent does choice of college major moderate the influence of undermatching on student labor market outcomes (i.e. annual earnings from employment)?

Background and Context

Though this study views undermatching as a phenomenon that negatively influences students' educational attainment and social mobility in general, it does not contend that every student should attend the most selective college possible. Some students might choose nearby less selective colleges because of family obligations, strong ties to home, or regional preferences (MacAllum, Glover, Queen, & Riggs, 2007). It is also true that some students are better off at less selective colleges. First, economically disadvantaged students attending elite colleges might experience difficulties because of class-based differences, as well as feelings of discomfort, inadequacy, exclusion and alienation (Granfield, 1991; Aries & Seider, 2005). Besides these challenges, they might also face problems of inadequate financial resources and cultural capital to adjust to their new circumstances (Aries & Seider, 2005). In contrast, those students at less selective colleges might find it easier to fit in with their new surroundings. Second, high-achieving students attending less selective colleges potentially have higher chances to get into honors programs, and have more interaction with faculty members (Fosnacht, 2015). In some sense, this is the "big frog in a small pond" phenomenon. In short, it is true that not every student should attend the most selective college to which they can gain admission.

However, even though there might be good reasons to choose a less selective college, it definitely should not be the norm (Bowen et al., 2009). As Bowen et al. put it,

> …a student should be made aware of the full range of higher educational opportunities available to someone with his or her credentials and then encouraged to reach for the

> most challenging opportunity that is a realistic option for the student. There may be compelling reasons for choosing what may be a "safer" of "more comfortable" option (to be nearer to home, for example), but such decisions should be made deliberately, after weighing all the pros and cons. (2009, pp 101)

To the contrary, previous research has found that under many circumstances, students undermatch because of inadequate information, insufficient planning, or simply inertia (Bowen, et al., 2009; Rodrick et al., 2008; Rodrick et al., 2009). These do not seem to be good reasons to undermatch. Moreover, choosing a "more comfortable" option is not necessarily more beneficial for students' educational attainment. Studies have shown that students graduate at higher rates if they attend a more selective institution (Horn, 2006; Gansemer-Topf & Schuh, 2006); particularly students from lower-income background, undermatching reduces six-year graduation rates (Goodman, Hurwitz, & Smith, 2014). In short, this study believes that the undermatching problem should not be ignored.

It has been revealed that less advantaged students, including low-income, first-generation, and underrepresented minority students, are populations that have higher tendency to undermatch. Students below the median socioeconomic status have undermatching rates almost twice as high as their higher socioeconomic status (SES) peers (Smith et al., 2013; Bowen et al., 2009). Moreover, low-SES students also tend to undermatch to a greater extent—the selectivity level of the college they attend is far lower than the ones they are qualified to attend. For example, it is not uncommon for a low-income high school graduate to choose to attend a nearby community college, even though they have a GPA of 3.6 and have taken two AP classes, which likely would easily have qualified them for admission to a moderately selective four-year university. Regarding ethnicity and race, though it is still controversial, some studies have found that minority students, especially Black and Latino students, are significantly less likely to enroll in a matched school (Roderick, Coca, & Nagaoka, 2011; Bowen et al., 2009). Students whose parents have a high school degree or less, and students living in rural areas also tend to undermatch at a significantly higher rate (Smith et al., 2013).

Reasons behind the undermatching phenomenon are multifaceted. One of the most obvious reasons is that some students are adopting unusual application strategies rather than following recommended processes. Students who tend to undermatch might only apply to a community college, or a non-selective four-year college that is close to home; other students may add one extremely selective college such as Harvard (Hoxby & Avery, 2013), to the list of applications they submit. In other words, their unique application behavior is ineffective in helping them find a matching school; to them, the application result is more like a lottery, instead of something they have a modicum of control over.

These unusual application behaviors are due to deeper causes. Less advantaged students often lack adequate information and support to navigate the college application process (Hoxby & Avery, 2012; Hoxby & Turner, 2013; McDonough,1997). Students who have early access to college information are more likely to choose from a wider pool of colleges (Rodrick et al., 2009). Specifically, selective institutions normally have deadlines earlier than non-selective four-year and two-year colleges, thus making early planning vital. However, underserved students often lack access to early information and planning, and consequently they often miss critical deadlines for submitting standardized tests, college applications and financial aid forms (Rodrick et al., 2008; 2009). In short, students who begin planning for college late generally have less chance of enrolling in a matching institution (Cochran & Coles, 2012).

Besides lacking early planning, underserved students also have limited sources of college information. Middle-class students can get college information from parents, college admission counselors, guidebooks, college rankings in newspapers (McDonough, Lising, Walpole, & Perez, 1998), the internet, or campus visits. By contrast, parents of low-SES and first generation students are less likely to know the way to navigate the processes of preparation, search, choice, and application (Choy, 2001; MacAllum et al., 2007). This is a major hurdle, as traditional-aged students are most influenced by their parents in the college application process, and such influence cannot be easily substituted by other individuals (Levine & Nidiffer 1996; Hossler & Gallagher, 1987; Hossler, Schmit, & Vesper, 1999; Cabrera & La Nasa, 2000).

Second, underserved students tend to rely largely on high school guidance counselors (Cabrera & La Nasa, 2000; MacAllum et al., 2007). However, at high schools serving primarily low-income students, counselors often have inadequate time to provide college advising, and students can hardly receive concrete and personalized attention (Rodrick et al., 2009). This occurs more frequently at schools serving predominantly Black and Hispanic students than at predominately white schools, which tend to have lower student-counselor ratios (Bryan, Moore-Thomas, Day-Vines, & Holcomb-McCoy, 2011). Moreover, sometimes high school counselors' misconceptions about financial aid and affordability also serve to discourage low-income students from attending a match institution. Besides, low-income students are often lacking necessary encouragement from teachers and counselors to properly evaluate their academic ability (McDonough & Calderone, 2006). As a result, less advantaged students often underestimate their academic abilities and eligibility for financial aid (Bowen et al., 2009; Cabrera & La Nasa, 2000). They also tend to overestimate the costs to attend college (MacAllum et al., 2007; Avery & Kane, 2004).

Research Design

This research utilizes the Educational Longitudinal Study of 2002 (ELS 2002) data from the National Center for Education Statistics (NCES), and supplementary data from American Community Survey (ACS), and the Integrated Postsecondary Education Data System (IPEDS). The ELS 2002 is a nationally representative, longitudinal study of 10th graders in 2002, and followed through their secondary and postsecondary years. In addition to the 2002 base year survey, it has three follow-up student surveys (2004, 2006, and 2012 respectively), as well as high school and college transcripts (restricted use). This longitudinal dataset includes variables such as students' demographic background, family background, high school experiences, college application and choices, college experience, and labor market outcomes. Therefore this dataset is ideal for answering the research questions listed above. ACS is a longitudinal household survey that gathers information about ancestry, educational attainment, income, language proficiency, migration, disability, employment, and housing characteristics at zip code level. IPEDS is a large-scale survey that collects institution-level data from postsecondary institutions in the United States by the NCES.

The operationalization of undermatching is based upon previous research which adopted a "non-parametric approach" (Belasco & Trivette, 2012, p.13), as employed by Belasco and Trivette (2012), Roderick et al. (2009), and Bowen et al. (2009). This method defines an eligibility criteria based on actual admission outcomes for institutions of each level of selectivity. Emphasizing a combination of admission standards, this method is specifically carried out as follows: For a particular combination of SAT (or converted ACT) score and high school GPA, if more than 90 percent of applicants were admitted into a specific level of selectivity, then all students who have the same or higher scores than the combination should get access to institutions of that selectivity level. Two sets of quantitative analysis were conducted for each research question, including descriptive analysis and Hierarchical Linear Modeling techniques.

Significance of study

As mentioned earlier, higher education has been expected to promote social mobility, to act as an equalizer, and to provide equal opportunities for every individual to succeed in life, regardless of social economic status, gender, racial background, and other circumstances (Haveman & Smeeding, 2006; Milburn, 2012). However, the discrepancy between the inspiration to succeed and the reality of limited opportunities has posed a huge problem: selective educational institutions permit fewer and fewer individuals to access ascending levels (Clark, 1960). Given the fact that one of the most observable benefits of investment in higher education is long-term earnings growth (Perna, 2006), the issue of who get access to a higher level and a more lucrative field should draw more attention.

When undermatching happens, high school graduates enter higher education institutions that are less selective than the ones their academic background allows them to get access. While this reality might undermine the social mobility goal of higher education to some extent, a limited number of studies have focused on the consequence of such phenomenon. Therefore, a thorough study on the influence of undermatching status on labor market outcomes allows for a better understanding of this issue. Moreover, considering the fact that major field of study might play a big role in deciding college graduates' income (Rumberger & Thomas, 1993; Thomas 2003; Pascarella &Terenizini, 2005), an understanding of how students' undermatching status and major field of study interact provides better insights to stakeholders and policymakers to support undermatched students.

Chapter 2: Literature Review

In this chapter I review the empirical research related to the undermatching phenomenon. More specifically, I organize my discussion into three sections: 1) college decision and application, 2) postsecondary undermatching and labor market outcomes, and 3) a review of current undermatching studies. The first two sections lay out a background discussion of undermatching phenomenon, and the third section trace the emergence and some of the early issues undermatching research addressed, and review current undermatching studies thereafter.

College Decision and Application

Bowen, Chingos and McPherson (2009) pointed out in their book that, students from low-income families are one of the major population groups that undermatching influences. Among students from family incomes in the lowest quartile, approximately sixty percent undermatched (Bowen, Chingos, & McPherson, 2009). By conducting analysis of four cohorts from 1972 to 2004, Bastedo and Jaquette (2011) found that, despite the fact that low-income students have made significant gains in pre-collegiate academic preparation, they still lag behind in college placement, especially in terms of enrolling in highly selective schools. Various mechanisms are behind the access problem for high-achieving low-income students, including information asymmetry, financial consideration, and so forth.

One ideological hypothesis of the phenomenon that only a small fraction of students in selective private colleges and universities are from families with the lowest incomes is that, low-income students are eligible for those selective schools but are simply excluded for the purpose of protecting children of the upper class (Winston & Hill, 2005). A general admission practice by elite colleges and universities is holistic review, which recruit students not solely based on academic merits, but also on other attributes that best meet institutional goals, such as athletic ability, artistic talent, race, gender, legacy status, economically disadvantaged background, and personal experience (Espenshade, Chung, & Walling, 2004; Fetter, 1997). While holistic review seeks to assess comprehensively the credentials applicants presented that facilitate successful college life and finally degree completion, students from wealthy families possess more economic, social and cultural capital to better develop and display those credentials. From this perspective, the holistic review admission process essentially is creating opportunities for the highly selective colleges to become the "bastions of privilege" (Bowen, Kurzweil, &Tobin, 2005) Schmidt, 2005). On the other hand, holistic admission might not be fully embraced in reality. A new study found that the personal backgrounds of admissions officers seems to have a major influence on how much they utilized holistic consideration to promote socioeconomic diversity (Bowman & Bastedo, 2016). Ironically, the system that originally intended to find talented students regardless of their background ends up identifying mostly students from privileged backgrounds (Avery, 2010).

Another explanation holds to a different argument: though the college admission system is impartial, there simply are too few low-income high-achieving students to qualify for selective colleges, as along their road from grade school to higher education they "slip through the cracks" (Wyner, Bridgeland, & DiIulio Jr, 2007, p.7). Economically ravaged neighborhoods, inadequate resources, weak family support, and problematic K-12 educational systems all collaborate to impair low-income students' educational opportunities and render them less successful than their peers from high-income families (Avery, 2010). Winston and Hill (2005) tried to answer the question of "are they out there"; that is, are there enough high-achieving low-income students in the pool for selective schools to enroll? To examine the distribution of family income of high-achieving students at a national level, analyzing the national population of high school test-takers in 2003, they found that the answer depends on how “high-achieving” is defined. In other words, the more relaxed the definition is, the more low-income students qualify. The authors found that an SAT or equivalent score of 1220 to 1420 as the criteria would support the statement that there are enough high-achieving low-income students for selective colleges to enroll. However as noted above, the admission process involves assessing not only SAT scores, but also other qualities, thus the conclusion based solely on SAT scores might be an over-simplification of the reality.

Despite these two different arguments, recent literature provides more insight in the recruiting and application process of low-income high-achieving student population. Hoxby and Avery (2012) pointed out that low-income high-achieving students are not applying in a recommended manner, and differ greatly with their peers from high-income families. Few low-income, high-achieving students have access to or follow the advice of expert counselors to apply to a few schools that are a "reach," four or more schools that are "par" or "match," and one of more schools that are "safe." The authors drew samples from students with college aptitude test scores that are in the top ten percent, and showed that more than forty percent of them apply to non-selective schools, and these schools are often local community colleges or local four-year institutions with low instructional spending per student and low graduation rates. Moreover, many apply only to a single non-selective college or to a single non-selective college and one additional college, which is often only weakly selective. Eight percent of low-income, high-achieving students apply in a manner that is somewhat close to what is recommended. And the remaining thirty-nine percent of low-income, high-achieving students use “unusual” application strategies, such as the combination of one local non-selective college and one extremely selective and famous college such as Harvard, or only applying to one public college within their state that is selective but is much less selective than the state's flagship university.

Such distinctive application behavior among low-income students is due to several reasons, with the most important one being, high achievers from low-income families may lack adequate information and support in navigating the college admission process (Avery, 2010). Unlike their middle-class peers, less-advantaged students can hardly get college information from parents, college admission counselors, guidebooks, college rankings in newspapers (McDonough, Lising, Walpole, & Perez, 1998), internet, or campus visits.

Regardless of SES or racial background, for traditional-aged students, parents exert the strongest influence on their college decision-making processes. However, as parental support requires specific levels of social capital, cultural capital and financial resources (Paulsen, 1990), parents who have a college degree are personally familiar with the process of college search, decision and application, and are able to provide more tangible support to their children (Choy, 2001). By contrast, parents of low-SES and first-generation students have lower possibilities to know the way to navigate the processes of preparation, search, decision, and application (Choy, 2001; MacAllum, Glover, Queen, & Riggs, 2007). Moreover, low-income parents also tend to know less about financial aid programs and qualifications. Previous research has shown that first-generation and low-SES students tend to receive less support from their parents in applying and choosing college (Levine & Nidiffer 1996; Hossler & Gallagher, 1987; Hossler, Schmit, & Vesper, 1999; Cabrera & La Nasa, 2000).

Secondly, underserved students tend to rely largely on high school guidance counselors (Cabrera & La Nasa, 2000; MacAllum et al., 2007). However, Rodrick et al. (2009) has found that counselors at high schools serving primarily low-income students often have inadequate time to provide college advising, and students can barely receive concrete and personalized attention. At schools serving predominantly Black and Hispanic students, this occurs more frequently as they tend to have lower student-counselor ratios (Bryan, et al., 2011). Except for the problem of inadequate access to counselors, sometimes high school counselors' also misconceive financial aid and affordability information, and this might also discourage low-income students from attending a match institution. Moreover, low-income students often lack necessary encouragement from teachers and counselors to properly evaluate their academic ability (McDonough & Calderone, 2006). Consequently less advantaged students often underestimate their academic abilities and eligibility for financial aid (Bowen et al., 2009; Cabrera & La Nasa, 2000), and overestimate the costs to attend college.

Meanwhile, low-income students can hardly get access to other channels of information. According to MacAllum et al. (2007), though most colleges have their admission websites, less-advantaged students often do not have internet access at home. For those who can connect to the web and find information, sometimes they experience problem understanding it, especially complex information about costs financial aid. In addition, college admission websites can be incomplete and outdated, which makes understanding even harder (Tucciarone 2009). Another source that is seen as democratized knowledge about college, national newspapers college ranking, also does not help much with situation. McDonough, Lising, Walpole and Perez (1998) demonstrated that for nontraditional, low-income, first-generation college and commuting students, national rankings published by major newspapers did not have much influence on their college decisions, indicating the significant difference in information acquisition approaches between students of different socioeconomic status.

Less advantaged students also have different perceptions about affordability of college.

By comparing students in the Boston public schools and those in affluent suburban schools, Avery and Kane (2004) demonstrated that the most important differences between the two groups lies in their perceptions of the ability to finance college, and the actions they actually take. Regarding the finance problem, more than half of the Boston public school students were unsure whether they could find a way to pay for a public four-year college in Massachusetts and eleven percent held a pessimistic view, while among suburban students more than seventy-five percent are confident that they could afford college and only one percent believe they could not. Students from low-income families are often discouraged by the sticker price, though the price they would actually pay was only a small fraction of the sticker price (Winston & Hill, 2005).

Moreover, low-income students are more deterred by the complexity involved in the college admission and financial aid application process. About two-thirds of the Boston public school students at the beginning of the senior year in high school in the study reported that they planned to attend a four year college immediately after high school graduation, but less than 25% realized this plan, and some never submitted an application to a four-year college (Avery & Kane, 2004). Additionally, among the Boston public school students who stated they wanted to go to a four-year college and have a GPA of at least 3.0, nearly ten percent did not register for the SAT while nearly fifteen percent registered but did not complete the exam. More than twenty-five percent students with high scores completed the SAT but never submitted an application to a four-year college.

According to Avery and Kane (2004), public school students experience hurdles with the SAT, mainly because of three factors. One is that those students were having problems completing registration forms, or began but never completed, or completed but never mailed it. The second reason is that very often students in public high schools registered too close to the deadline and then are assigned to unfamiliar suburban test centers that might be a thirty-minute drive. Some were discouraged by the distance, some got lost on their way to the test locations, and some arrived at the correct site but did not bring valid picture IDs ore registration slips. And finally some Boston public high school students just made last-minute decisions that they would not take the test. Still, for students who successfully completed the SAT, it is reported that they experience substantial difficulty in writing application essays.

The situation that low-income high-achieving students face in applying to selective colleges is accompanied by the fact that college admission staff using common college admission approaches also experience difficulty approaching low-income high-achievers. Hoxby and Avery (2012) found that most high-achieving, low-income students do not apply to any selective colleges, and this is accompanied by the fact that they are unlikely to be discoverable by college admissions staff through traditional recruiting approaches. Five traditional approaches were analyzed by the authors. First, colleges send brochures to students on the mailing lists they purchase from the ACT or College Board, and those on the lists are students who meet specific criteria, such as achieving high scores on the college assessment exams. These brochures, however, do not differentiate between low-income students and high-income students or provide specific information about cost and financial aid that low-income students need in particular. Second, some selective colleges recruit qualified, low-income applicants through counselors from magnet high schools, who tend to practice relatively strict admissions. Third, some selective colleges rely on college access programs to recruit high-achieving low income applicants, and many of these access program activities are “self-selected”, and it is unknown whether they actively search for qualifying students or simply act as a pipeline. Fourth, college admission staff make recruitment trips to high schools to meet prospective students. However with the tremendous number of high schools in the United States, only a small portion are visited. To make recruitment trips more efficient, targets are often located in a place that students from several high schools can attend. And finally, colleges hold open-house activities to recruit applicants, so for low-income students probably only those who happen to live nearby will have a chance to make the trip. In short, utilizing these regular recruiting activities, it is challenging for colleges to reach the high-achieving, low-income students more effectively and efficiently.

College Major Choice

Compared to the abundant studies and reports on college decisions and applications, relatively few have focused on the choice of college major field of studies. Previous research has considered demographic background, the influence of individual subjective factors such as interest in the subject and self-efficacy, family background such as parental income and occupations, and economic perspective, such as expected earnings. Specifically, some studies also examined the intersection between the factors mentioned above and gender/race.

Early studies have noticed the gender pattern in college major choosing. It is reported that male students tend to choose business, engineering, chemistry, and physics, while female students typically are more interested in language, literature, arts, nursing and teaching (Polachek, 1978; Jacobs, 1986; Jacobs, 1995; Davies & Guppy, 1997). From another perspective, women pay less attention to economic benefits when choosing a field (Davies & Guppy, 1997; Jacobs 1995; Hecker, 1995; Marini & Fan, 1997). Studies have also documented the racial differences in college major choices. Compared to White and Asian peers, Hispanic and African American students are less concentrated in science majors (Simpson, 2001; Berryman, 1983; Maple & Stage, 1991; Mullen, 2001; Powell, 1990).

Self-efficacy is found to be influential as well. Individual estimation of their ability to succeed in particular majors influences their major choice. In other words, if perceived ability is equalized, the number of students choosing economics would increase, and the number majoring in humanities would decrease (Arcidiacono, Hotza, &Kang, 2012). Specifically, researchers have long noticed that mathematics self-efficacy is significantly related with choosing science-based college majors (Betz & Hackett, 1983; Hackett, 1985; Lent, Brown, & Hackett, 1994; Maple & Stage, 1991; Ware & Lee, 1988). Betz and Hackett (1983) found that mathematics self-efficacy expectations are as important as math ability per se in influencing choosing science-based college majors. Hackett (1985) used path analysis to confirm the hypothesis, that mathematics related self-efficacy moderates the effects of gender and mathematical preparation and achievement on math relatedness of college major choice. On the other hand, Astin (1993) found that self-rated writing skills were strongly related with major choice, and that students with high self-rated writing skills were less likely to major in engineering and math.

A large body of research also suggests that personality is highly correlated with choosing a particular college major. For example, Astin (1993) found that students interested in social activism were more likely to choose social sciences and education as majors, and those who rated high on the artistic scale were more likely to major in arts, music and theater. Later literature took a step further and investigated person-environment fit, applying Holland (1985)'s theory of careers. The tenet of the theory is that the choice of a vocation or a college major is an expression of personality, and most people can be categorized into one of the six personality types, including realistic, investigative, artistic, social, enterprising, and conventional (Smart, Feldman, & Ethington, 2006). Students choose academic environment that is congruent with their personality types, and such congruence is related to college success through influencing students' educational stability, satisfaction and achievement (Smart, Feldman, & Ethington, 2000; Feldman, Smart, & Ethington, 1999; Porter & Umbach, 2006).

Family background factors that have been proven to be influential include family income and parental occupations. One theory in early studies assumes that socioeconomically advantaged parents would secure the best position for their children in the education system, both in terms of level of education and field of study (Gamoran & Mare 1989; Hallinan, 1992). Under such assumption, students from higher SES families would choose college fields that predict better economic returns. However, other research works have disagreed with such assumption, and found out that students from wealthier families place more emphasis on rewards other than financial benefits (Kohn & Schooler, 1983), and working-class students are more likely to treat college education as an approach to move upward (Davies & Guppy, 1997).

Specifically, some researchers have also explored the intersection effect between gender and family SES on students' major choice. Green (1992) found that among students from wealthy families, more male major in business than female. Trusty et al. (2000) indicated that women's major choice were more influenced by their socioeconomic status than man, and women from high SES background were more likely to choose science and business majors than other women. However, Ma (2009)'s research put forward a different view. Her research found that women and men from lower SES backgrounds are equally likely to choose lucrative college major; however, women from high SES families incline toward social science and humanities, while men from high SES families tend to choose business and life/health majors. Therefore, previous studies have not reached common conclusion regarding the differential effects of SES on college major choice for different gender.

Besides these non-pecuniary factors mentioned above, some research have also considered the impact of pecuniary factors, such as expected labor market earnings. To date, there are no conclusive answers to whether pecuniary factors are more important than other factors. An early study by Berger (1988) showed that holding family background characteristics constant, when choosing college majors, students are more influenced by the expected flow of future earnings than expected beginning earnings. Arcidiacono et al. (2012) found that expected earnings in different majors is an important factor when choosing college major. Specifically, their model also demonstrated that students' expected earning differences of themselves and peers in different majors also influence students' major choice. On the other hand, study by Beffy, Fougere and Maurel (2012) suggested that the impact of expected earnings on major choices is statistically significant; however, such impact is quantitatively small. Their conclusion were in accord with Staniec (2004), that expected returns have no significant effect on the probability of choosing different major categories.

Postsecondary Education and Labor Market Outcomes

It has been widely recognized that higher education is one of the best investments an individual can make. Sources have shown that attaining a higher educational degree implies greater success in the labor market. The 2012 NCES report indicated that among individuals of twenty five and over, the median earnings of those who graduated from high school but did not attend college is less than $30,000 per year, while those whose highest level of educational attainment was a bachelor's degree earned around $50,000 per year. Those with advanced degrees have even higher yearly earnings, with those holding a master's degree at approximately $60,000, a doctorate degree at $82,000, and a professional degree at $97,200.

However, with the current situation that sixty-six percent of high school graduates enroll in postsecondary institutions (Bureau of Labor Statistics, 2014), the hierarchical structure within higher education system increasingly exerts a crucial influence on the distribution of life chances. There are two hierarchical structures within the higher education system: selectivity, and majors (in terms of economic returns).

Institutional selectivity and labor market outcomes

The fact that a growing population is entering the higher education system, and that an increasing number of bachelor's degrees are conferred has rendered college degrees to be less of a guarantee for labor market success. Greater focus now must be placed upon where individuals actually go to college, instead of simply whether one went to college or not. Research on the relationship between postsecondary schooling and earnings is extensive, with the majority focusing on institutional characteristics, such as selectivity or college quality.

Earlier studies showed that college selectivity has a significant positive effect on graduates' earnings. James, Alsalam, Conaty and To (1989) used the National Longitudinal Study of the High School Class of 1972 (NLS72), and examined the effect of college selectivity on earnings for male college graduates. Their analysis showed that graduates from private Eastern universities, which they considered to be the most elite and selective institutions, earned wages that were eight percent higher than those from public institutions. Moreover, the average SAT score of the freshman class also significantly affected earnings: for every 100-point increase in SAT average score, annual earnings raise about three percent.

Fox (1993) used the 1980 High School and Beyond (HSB) survey of high school seniors and follow-up surveys in 1982, 1984 and 1986 to test the relationship between hourly wage and selectivity of the college attended. He constructed a college selectivity measure from Barrons'

Profiles of American Colleges, incorporating incoming students' class rank, high school GPA, average SAT score, as well as acceptance ratio, and grouped colleges into six levels of selectivity. After controlling for individual and family characteristics, Fox (1993) found that graduating from a highly selective college generated a wage premium of 13 percent.

Brewer, Eide, and Ehrenberg (1999) utilized both NLS-72 and HSB, and also adopted Barrons' Profiles of American Colleges selectivity categorization. They further collapse the original six categories into three, and added another dimension of institutional control. In this way they create six categories of institutions: elite private, elite public, middle private, middle public, bottom private, bottom public. Their results show that in the 1982 cohort, students who earned a degree from an elite private college, as compared to those who graduated from the bottom pubic institutions, earned 37 percent higher wages on average.

Utilizing the National Longitudinal Survey of Youth, which surveyed individuals from 1979 to 1996, Monks (2000) examines how college characteristics influence returns to individual. He controlled for institutional characteristics such as selectivity, control, and college type, as well as individual traits such as actual academic ability, and work experience. His finding indicates that graduates from graduate-degree granting research institutions earned approximately 14 percent more than their peers from liberal arts colleges. Regarding institutional selectivity, graduates from highly competitive institutions have earnings 15 percent more than competitive college graduates, while the latter earn approximately 5 percent more than graduates from non or less competitive institutions.

However, recent studies have found contradicting results, showing a weak or insignificant impact of college selectivity. Dale and Krueger (2002) recognize the selection problem, which refers to the possibility that selective institutions admit student based in part on characteristics that also determine their earnings capacity. Researchers may not account for these characteristics when they conduct studies focused on predicting wages; therefore separating the effect of graduating from a selective college from individual pre-college characteristics is difficult. To avoid the influence of unobserved characteristics on college admission, Dale and Krueger (2002) use College and Beyond (C&B) Data Set and match students who were either admitted to or rejected by institutions with similar qualities. They found that the earnings of students who attended more selective colleges do not differ significantly from those who were accepted but attended less selective colleges. However they find that the average tuition charged by the college significantly impacts graduates' earnings. Unfortunately, the sample in their study was limited to enrollees at highly selective colleges, thus reducing the generalizability of their findings to institutions of different quality (Long, 2008).

Recognizing the problem of making linearity assumptions in previous literature, Black and Smith (2004) estimated the impact of college quality on wages using propensity score matching methods. They used the National Longitudinal Survey of Youth 1979 cohort, and reach different results when utilizing different methods. When using an ordinary least squares (OLS) methods, they found positive significant effects of college quality; however when they used a propensity score matching technique, the significant effects disappeared. Similarly, Long (2008) utilized the National Education Longitudinal Study (NELS) data, and found that the OLS estimates tends to indicate positive and significant impact of college quality, while non-OLS specifications are more likely to show insignificant results. However recent studies also reached different conclusions. For example, adopting a regression discontinuity approach, Hoekstra (2009) demonstrated that students attending a state flagship earn twenty percent more than those who attended less selective colleges.

Just as Dale and Krueger (2011) pointed out, drawing reliable conclusions can be difficult, because of unobserved characteristics such as ambition, organizational skills, and time management skills. These unobserved characteristics can potentially influence students' enrollment in college as well as success in the labor market (Gunderson & Oreopoulos, 2010). What these findings point to is the need for more advanced and particularized studies of college attendance (including college decision) and labor market outcomes.

Academic major and labor market outcomes

The labor market outcomes of college graduates have been found to be substantially influenced by two factors: college selectivity and college major. The influence of college selectivity on earnings has been discussed in the previous section, and generally speaking, graduating from a more selective and high quality institution predicts significantly higher wages (Thomas & Zhang, 2005). However, an equally influential factor is identified as the major field of study students choose. As a factor that has long been recognized, major field of study exerts great influence on college graduates' labor market outcomes (Rumberger & Thomas, 1993; Thomas 2003). On the whole, more lucrative majors "have a relatively specific and well-defined body of content knowledge, and focus on methods of inquiry requiring high levels of quantitative or scientific skills" (Wolniak, Seifert, Reed & Pascarella, 2008, p.125). Specifically, college graduates who major in Science, Technology, Engineering, and Math (STEM) tend to have higher salaries than students who major in other disciplines. According to Pascarella and Terenizini (2005), majors directly account for twenty-five to thirty-five percent of the earnings effects of higher education. Also, in a 2012 report, the annual income for graduates in engineering and computers and mathematics are $55,000 and $46,000 respectively, while the figure for humanities and liberal arts are $31,000 (Carnevale, Cheah, & Strohl, 2012).

Adding to this general research results that STEM graduates earn more, recent studies shed some light upon the interaction between college major and other factors in influencing earnings. Using data from colleges in the Appalachian Region in 2001, Wolniak et al. (2008) find that the earning effect of background characteristics such as gender, parents' education, and family income are not uniformly moderated by education attainment across majors. For example, among Health Sciences and Social Sciences majors the moderating effect of educational attainment is significant, while among STEM majors the effect is inconsequential. Zhang and Thomas (2005) consider college quality when estimating the effect of academic major on earnings, and find that there are substantial differences. Business and social sciences graduates' earnings differ greatly by college quality; however for engineering majors, college quality does not significantly determine their earnings. Surprisingly for health majors, college quality negatively influences earnings; the authors provided an explanation, that health industry is heavily technical skill-based, and high-quality schools often provide less vocationally oriented programs. Similarly, Eide, Hilmer and Showalter (2015) suggest that major-specific earnings varies, based on college selectivity. They also find that the strongest differences exist among business majors, and the weakest among science majors.

Current Undermatching Research

Emergence of undermatching study

Academic leaders and policymakers have long been focusing on the equality of opportunity to attend higher education institutions. Today, in addition to the question as to whether a high school graduate can attend a college, another relevant question concerns what college one attends. The American higher education system is highly stratified, with the bottom level with the least selective institutions being “egalitarian”, such as community colleges which practice open admission, and the upper level being “elite”, such as private liberal art colleges which normally are considered to be more prestigious, and practice more strict admission policy that is based on meritocratic criteria.

Scholars in sociology of education have identified three types of equality of opportunity. According to Hearn (1991), the first one is “equity of condition” (Hearn, 1991, p.160), that all students should receive similar quality of schooling regardless of their background. The second definition of equality adds a meritocratic component, requiring equity of condition among students of equal academic achievement, and educational aspirations. Accordingly, this definition suggests that the best students should attend the best college. In contrast, the third definition emphasizes outcome equality, which aims to narrow the gap among students with different academic achievement. This leads to the practice of providing the best schooling to the least prepared, in other words, this equality is “redemptive” (Hearn, 1991, p.160).

The argument for the college undermatching issue is therefore derived from the principle of meritocratic equity of condition, and its underlying ideology is that the student-college matching should be based on their prior academic achievement. Numerous studies subscribe to this view and have been exploring the issue of college access, especially in terms of low-SES, high-achieving students (Wyner, Bridgeland, & DiIulio, 2007; Stage & Hamrick, 2003; Hoxby & Turner, 2013). These studies recognized the problem that some students’ hard work during high school did not pay off-- they failed to enroll in a college that their academic background presumably enabled them to attend. However, these studies do not clearly explore undermatching to the full extent (Smith, pender, and Howell, 2013). Important questions began to arise around 2006 through a series of research reports by Roderick and colleagues in which they started to operationalize the notion of college undermatching (Roderick, Nagaoka, Allensworth, Coca, Correa, & Stoker, 2006; Allensworth, 2006; Roderick, Nagaoka, Coca, & Moeller, 2008; Roderick, Nagaoka, Coca, & Moeller, 2009; Roderick, Coca, & Nagaoka, 2011).

The above mentioned series of research reports are the products of a multi-year project *From High School to the Future,* executed by the Consortium on Chicago School Research. In their quantitative project they tracked successive cohorts of Chicago Public School (CPS) high school graduates, and collected data regarding the relationship among high school preparation, college decisions and higher education outcomes (Roderick et al., 2006). In their first report in 2006, in order to answer the question of whether CPS students' access to college and especially four-year college is constrained by their qualification, they estimated the percentage of CPS graduates whose academic background suggest that they should have access to at least a four-year college. Their analysis also shows that many CPS high schools have more graduates who are qualified to attend a somewhat selective or more selective college than those who actually enroll. These results suggested that more efforts should be made to translate qualifications into enrollment so that more students who are actually qualified for more selective four-year colleges attend them.

In their 2008 report, Roderick and colleagues focused on the pipeline from high school to college, and placed much emphasize on understanding the undermatching problem. They examined whether CPS students who aspire to attend four-year colleges were effectively engaging in the college search and application process. From aspiration to enrollment, there are three stages that high school graduates must go through: planning to attend a college in the fall, applying to a college, then being accepted and choosing to matriculate into a college. One of their key findings is that, after going through the above-mentioned process, only one-third of

CPS students who aspired to achieve a four-year degree attended a college that matches their academic qualifications. Especially, among those who possess the highest qualification, only 38% enrolled in a matching college, and an equal percentage enrolled in a college far below their apparent level of achievement. Moreover, undermatching is an issue for students at all levels of qualifications, instead of only for students of the strongest academic background (Roderick et al., 2008).

The 2009 report further explores the undermatching problem of academically advanced students. Three groups of students were examined, including graduates from CPS's selective enrollment high schools, graduates from International Baccalaureate (IB) programs, and graduates who had taken Advance Placement (AP) and honors courses. They chose these three groups of students as they possess the strong qualifications that are required for enrolling in more selective colleges. The authors found that these high-achieving students did not necessarily come from more advantaged communities or families, and faced the same or even more challenges navigating college application. For example, some highly selective colleges have more complicated and specialized application procedures, like the consideration of "legacy" admits – those students who are the children of alumni, which create further barriers for high-achieving students. As a result, fewer than half of these students enroll in colleges that match their academic qualifications; moreover, one-fifth of those students did not even apply to a four-year college (Roderick et al., 2009). These series of reports helped to promote what proved to be a growing interest in college undermatching; since then subsequent research burgeoned.

Review of current undermatching studies

Some key literature has shed light upon the nature and complexity of undermatching in postsecondary education. Regarding the scope and general trend of postsecondary matching, existing studies have produced distinct estimates using different undermatching definitions and statistical techniques. For example, Smith, Pender and Howell (2013), used the Education Longitudinal Study of 2002 (ELS:2002) dataset to examine the cross-cohort differences in the extent of undermatching, by comparing the 1992 cohort with the 2004 cohort nationwide, and found out that the 2004 cohort was eight percent less likely to undermatch than the 1992 cohort. Moreover, while the undermatching rate for all students showed a decline, the declines are larger for lower-SES students, pointing to a closing of the SES gap in undermatching. The authors attributed such a decrease to a rise in students enrolling in nonselective and two-year colleges. Belasco and Trivette (2015)'s research, also using the ELS 2002 dataset, indicated undermatching as a less pervasive problem, showing a 28% undermatching rate nationally. This result, as Belasco and Trivette themselves put it, was due to a different definition of undermatching, as well as a more sophisticated techniques adopted than Smith et al (2013).

Research has also identified predictors for undermatching. Aside from the aforementioned SES impact, factors such as racial background, student attitude, and high school context have been shown to have an effect on undermatching. But the research findings offer some divergent results. For example, there exist contradictory results regarding racial background. Bowen et al (2009) and Roderick, Coca and Nagaoka (2011) found respectively that African Americans and Hispanic students are more likely to undermatch, while Belasco and Trivette (2015), as well as Smith et al. (2013) indicated that African-American, Asian and Hispanic students have a lower probability to undermatch, and especially African-American are significantly less likely to undermatch by selectivity. Students' preferences while making a postsecondary decision is also a significant factor; for example, students who emphasize low college tuition and living at home have a higher likelihood to undermatch, while students attaching high importance to social life in college were less likely to undermatch (Belasco and Trivette, 2015).

Students' admission-related activities are also shown to have an influence on lowering the likelihood of undermatching, such as visiting a school counselor, completing a FAFSA, and submitting a higher number of applications (Belasco & Trivette, 2015). Moreover, high school context matters in the process of translating students' aspirations into enrolling in an institution that matches their academic capability: Roderick, Coca and Nagaoka (2011) concluded that if urban high schools lacks an organizational culture that can effectively guide students through the college application process, then students' qualifications and college aspirations will not necessarily lead to four-year college enrollment, thus potentially promoting undermatching. Similarly, research by Belasco and Trivette (2015) showed that students attending schools with a strong college-going culture have a lower likelihood to undermatch.

As a research topic that caught the attention of the educational research community approximately a decade ago, there exists relatively few article that focus on the influence of postsecondary undermatching on students' outcome. Research has shown that if students choose to attend a college the selectivity of which matches their measured academic ability, there is higher chance that they will complete a degree, and this is true for students of all academic ability levels (Light & Strayer, 2000; Hughes, 2013). In theory, it is not necessarily the case that undermatched students are less likely to graduate (Smith et al., 2013); one might consider the frog-pond effect, that given similar ability level, high-performing students at academically inferior schools may develop a more favorable self-concept than low-performing students at superior schools (Alicke, Zell, & Bloom, 2010), and such psychological advantage should ultimately be conducive to degree attainment. Fosnacht (2015) also found that undermatched students interact more frequently with faculty and are more engaged in collaborative learning activities.

However, a variety of reasons for lower graduation rates among undermatched students have been revealed. Peers, student body makeup, and campus culture could influence students' learning habits, and degree aspirations (Smith et al., 2013). Also, less selective institutions typically tend to have smaller budgets and a lower expenditure per student (Hoxby, 2009), which might result in less academic and non-academic support, and ultimately lower graduation rates. Moreover, Fosnacht (2015) reported that students who attended less selective institutions feel less academically challenged, perceive less benefit from attending college, and are less satisfied with their experience than students who attend institutions matching their prior educational achievement.

To date, one research has explicitly focused on the labor market outcomes of undermatched students. Hughes (2013) followed the method of Dale and Krueger (2002), which used the matched applicant model that make comparison between students accepted to and rejected by the same institutional types. By utilizing the Bureau of Labor Statistics' 1997 National Longitudinal Survey of Youth data, Hughes' (2013) result suggested that undermatching and overmatching only weakly influence students' earnings. However this research does not take into account the factor of major field of study.

Conceptual Framework

As this research aims to explore factors influencing undermatching, and the labor market outcomes of undermatching, two sets of conceptual frameworks are utilized to guide each part of the study. Research question 1 (how do students' background characteristics influence their chances of undermatching) and question 2 (how do students' background characteristics, especially undermatching status, influence their probability of choosing STEM majors) mainly uses Perna (2006)'s college access and choice model, while also integrating some components of Iloh (2018) college-going decision model. Research question 3 (to what extent does undermatching influence students' labor market outcome and how does college major moderate such influence) is based on human capital theory.

College access and choice model

A comprehensive student success framework by Perna (2006) was developed to synthesize the disagreement among higher education researchers about college choice, access and success, especially regarding the unequal distribution of academic and financial aid resources. The Perna (2006) student success framework includes four key phases of student success, namely, college readiness, college enrollment, college achievement, and post-college attainment. College readiness refers mainly to the status before higher education enrollment, including educational aspirations and academic preparation. College enrollment has two components, college access, and college choice. College achievement includes academic performance and persistence, while also examining transfer. Lastly, post-college attainment primarily focuses on educational attainment, and income. Perna (2006) also further developed a college access and choice model using a similar framework, and the review below provides a detailed description of this model.

Previous research reached different conclusions about what caused the college access gap, such as lacking appropriate academic preparation (Ellwood & Kane, 2000; Perna, 2005), inadequate financial aid programs (Fitzgerald, 2004; St. John, 2003), and limited information about academic and financial resources (Kane, 2010). Therefore, Perna (2006) proposed a conceptual model that integrates several different disciplinary approaches and perspectives, such as sociological, economic and cultural capital, to better understand the problem of student college choice.

In Perna (2006)'s conceptual model, individual's college choice is determined by four contextual layers, including the individual habitus, the school and community context, the higher education context, and the broader social, economic, and policy context. The individual habitus layer, as the center of this model, combines human capital investment theory, cultural capital, and social capital. Simply stated, individual's college choice is based on their comparison between the expected benefits and the expected costs. In addition, their expected benefits and costs are also influenced by their economic capital (eg. family income, financial aid), cultural capital (eg. cultural knowledge, value of college attainment), and social capital (eg. information about college, resources with college processes), along with their demographic background (eg. gender, race).

The core individual habitus layer is then nested within the second layer—school and community context. School and community context is constructed, based on the assumption that individual's behaviors are best understood within a broader context (Bourdieu, 1986; Lin, 2001). Several aspects of high school contexts can influence student college choice, such as the availability of resources, type of resources, and structural support and barriers. Examples of such resources include carefully designed school guidance process (Schneider & Stevenson, 1999), availability of staff and counselors that are familiar with curricular requirements and paths (McDonough,1997), and the amount of college counseling student can get access to, and supplemental programs and services (González, Stone, & Jovel, 2003).

In addition to individual habitus layer, and high school and community layer, the higher education context also influences student college choice. Factors that have been identified include institutional recruitment and marketing (Chapman, 1981), geographic location of institution (McDonough, Antonio, & Trent, 1997; Perna, 2000), and competition for access to elite institutions (McDonough, 1997). Specifically, geographic location reflects different regional tradition and philosophy towards education, represents different system and size of higher education (Perna & Titus, 2004).

The final layer is the social, economic, and policy layer. In particular, the social context of a region includes demographic characteristic, and examples are percentage of population with bachelor's degree, poverty rate, and percentage of Hispanic population (St. John, 2003). The economic context refers to characteristics of labor market. The most direct index is the unemployment rates, while other social changes have also been found to be influential. An example of such social change could be the industrial paradigm shift, when good-paying jobs started to require education beyond high school (Bettis, 1996). Policy context includes policies and structures that impact higher education enrollment. In addition to higher education financial aid and tuition policies (Titus, 2006; John, Hu, & Weber, 2001), K-12 educational policies also effect college enrollment: K-12 curricular requirements and assessments would need to align with college enrollment to ensure a smooth transition from high school to college (Venezia, Kirst, & Antonio, 2003). And lastly, affirmative action policies have been found to impact college enrollment behavior of students of different ethnic background (Long, 2004; Horn & Flores, 2003; Hinrich, 2012; Backes, 2012).

The Iloh model of college-going decisions and trajectories

The Iloh (2018) model of college-going decisions and trajectories was developed to challenge the dominant college choice model, and the idea of "choice" itself. One of the most recognized college choice models, the three-stage linear model developed by Hossler and

Gallagher (1987), depicts the sequential procedure of traditional high school student's college-going process. In short, firstly high school students are predisposed to enter postsecondary education, then they gather information about possible higher education options, and lastly, they make a choice about which college to attend.

The limitation of the Hossler and Gallagher (1987) model, however, lies in the fact that college-going patterns are increasingly becoming diversified. Instead of attending college as a one-time four-year experience, high school students nowadays might attend college at a different time point of life, attend more than one college, or leave for some time and return later. Therefore, these students might utilize a different approach to conceptualize college opportunity, and evaluating their own chances of access to colleges, and to certain level of colleges. On the other hand, because of the limited resources and different pattern of college opportunity conceptualization, disadvantaged students do not have the option to "choose" from abundant options. The phrase "choice" therefore did not capture the realities of complicated situations faced by nonconventional students.

Therefore, the new Iloh (2018) model takes an ecological perspective, identified three dimensions that determine individual's college decisions, and highlighted the interactions among these three dimensions: information, time, and opportunity. The relationship among these three dimensions are nonlinear and codependent, and could be applied to different time points in life (Iloh, 2018).

The information component focuses on the quality, quantity, and delivery of higher education related information. This information could be general and objective statement of college information, or subjective and customized suggestions or warnings. General and objective college information are primarily college facts, while subjective and customized suggestions or warnings contain the source's evaluation of fit of specific institutions, for example, a specific college does not offer much support for low-income students, or faculty of color at an institution often leave so be cautious (Iloh, 2018; Iloh, 2019). The delivery of information is related to the approach information is communicated, or the who and how of the message (Iloh, 2018).

Time is another component, and helps understanding the context of individuals at different stages in life. Based on the Bronfrenbrenner and Morris (1998) bio-ecology model, Iloh (2018) includes three ecological forms of time, micro-time, meso-time, and macro-time. Micro-time refers to the immediate events that individuals experience, such as scholarship application being rejected. Macro-time is related to the broad circumstances that changes over time, for example, a state's higher education legislation and regulation change over a period of time span. Lastly, meso-time lies in between, refers to the consistency and interaction between individual's activities and environment. One example of meso-time could be someone driving past a billboard for a particular college daily on their way to work (Iloh, 2018), or someone continuingly saving for several years to finance college (Iloh, 2019).

The third component is opportunity, either factual opportunity, or perceived opportunity. On one hand, different background aspects, including identity, life experiences, spatial, financial, parental contexts, and broader high school and community contexts, all influence the factual opportunities individuals could grasp. On the other hand, these same sets of factors could also influence their perceived probability of access and fit of particular higher education institutions. As a result, despite the fact that a plethora of higher education options exists, many do not consider them to be available to them.

In summary, this new model sheds some light on the college-going process, and provides a new perspective to complement the Perna (2006) student success model mentioned earlier. The current study adapts and utilizes the Perna framework for college decision and major choice, while also integrating the *information* and *perceived opportunity* components from the Iloh (2018; 2019) model. Specifically, the framework for this study includes three layers: individual layer, high school layer, and community layer. Detailed conceptual framework will be specified at the end of this chapter.

Human Capital Theory

The concept of human capital dates back to Smith (1776): "The acquisition of …talents, by education, study, or apprenticeship, always costs a real expense, which is a capital fixed and realized. Those talents, as they make a part of his fortune, so do they likewise of that of the society to which he belongs." (p.217). Becker (1975) and Mincer (1958) further extended the human capital concept, and developed a theory that explains wage differences of employees. According to Becker (1975) and Mincer (1957), the differences among personal incomes are due to the different amount of human capital employees invest and own, including schooling, training, and other forms of education.

In particular, human capital differs among individuals, and sources of variation include innate ability, schooling and school quality, non-schooling investments, training, and pre-labor market influences (Acemoglu & Autor, 2011). Innate ability refers to the biological characteristics, such as some components of Intelligence quotient (IQ), and disposition. Therefore, suppose individuals receive the same education, training, and other investments, it is still highly possible that these individuals possess different amounts of human capital.

Schooling, or formal education, has been widely studied, since it is one of the most observable investments in human capital. Human capital theory emphasizes that formal education increases efficiency and therefore can lead to greater economic productivity (Becker, 1994; Kern, 2009). However, it is worth noting that the explanatory power of schooling (as shown by the R^2 in regression models) in the general human capital model is relatively small, and this suggests that there are other important factors of investment in human capital. Still, schooling serves as an important component of analysis, as factors that influences schooling could also influence non-schooling investments (Acemoglu & Autor, 2011). . One specific aspect of schooling is the school quality, such as school attendance rates, teacher/student ratio, and relative teacher pay. School quality has also been found to have an effect on the rate of return (Card & Crueger, 1992; Hanushek & Woessmann, 2016).

Non-schooling investments are investments in other components of human capital that individuals choose to make, and sometimes are seen as unobserved skills. For instance, one might choose to work harder, or study especially for some areas, or develop different personality and non-cognitive skills. These unobserved skills are seen as equally important factors that influence the distribution of wages, and possible wage changes (Green & Riddell, 2003; Chay & Lee, 2000). One concerning problem here is that, it is relatively hard to obtain appropriate data on this component of human capital. Some approaches have been developed to cope with this problem, such as developing proxy for unobserved ability, making general assumption that the impacts of unobserved ability are constant during the whole life circle, or utilizing specifications of an education production functions that controls for time-varying unobserved ability (Ding & Lehrer, 2014).

Training is another source of education, and is normally acquired after schooling. The subject of training is normally related to a specific set of skills that can be directly used in a job or industry. On one hand, training is similar to schooling that individuals can control for the amount of investment; on the other hand, training is different from schooling that employers jointly invest in the training of individuals, because the result of training directly matches the need of employers.

The last component, named “pre-labor market influences” by Acemoglu and Autor (2011), refers to the peer group influence before individuals enter the labor market. While such concept is close to the notion of social capital, Acemoglu and Autor (2011) emphasized an element of investment. One example of the pre-labor market influences is that, when parents are making decisions about where to purchase a house, their decision is also related to what kind of pre-labor market influences their children will receive.

However, in real world labor market, the relationship between human capital and the salary individuals receive is not that linear and straightforward. Three typical caveats of the human capital model exist, including compensating differentials, labor market imperfections, and taste-based discrimination (Acemoglu &Autor, 2011).

Compensating differentials, or equalizing differences, refers to the situation that the wage rate and pleasantness/unpleasantness of a particular job are balanced. This implies two different scenarios. One scenario is that when the employees are receiving additional compensation for undesirable attributes of a specific job, such as higher effort requirement, and less enjoyable working environment. An opposite scenario is that when employees are receiving fewer wages, in exchange for more desirable working environment, or other special non-monetary benefits (Rosen, 1986).

Another source of caveat is the labor market imperfections. One example of labor market imperfection is that an employee's wage is connected to the productivity of a job, and jobs vary in productivity. As employees are matched to jobs of different productivity, this could result in employees that possess the same amount of human capital receive different wages (Leibenstein, 1957; Stiglitz, 1976). The last caveat here is taste-based discrimination, when employees are paid less wage, because of their gender, race, or other demographic characteristics (Charles & Guryan, 2009; Carlsson & Rooth, 2012).

Despite these caveats, human capital theory is still the most powerful framework that contributes to our understanding of education and productivity. The current study about the labor market outcome of undermatched population is based on the human capital theory framework, and detailed variables can be seen in chapter 3.

Adapted conceptual framework

For the first and second research question, combining Perna (2006) college access and choice model, and integrating components from Iloh (2018)'s alternative model of college-going decisions, the adapted framework is shown in Figure 2.1. The conceptual framework of the third research question is based on the human capital theory, and could be seen in Figure 2.2

Figure 2.1 Adapted Conceptual Framework for Research Question 1 and 2

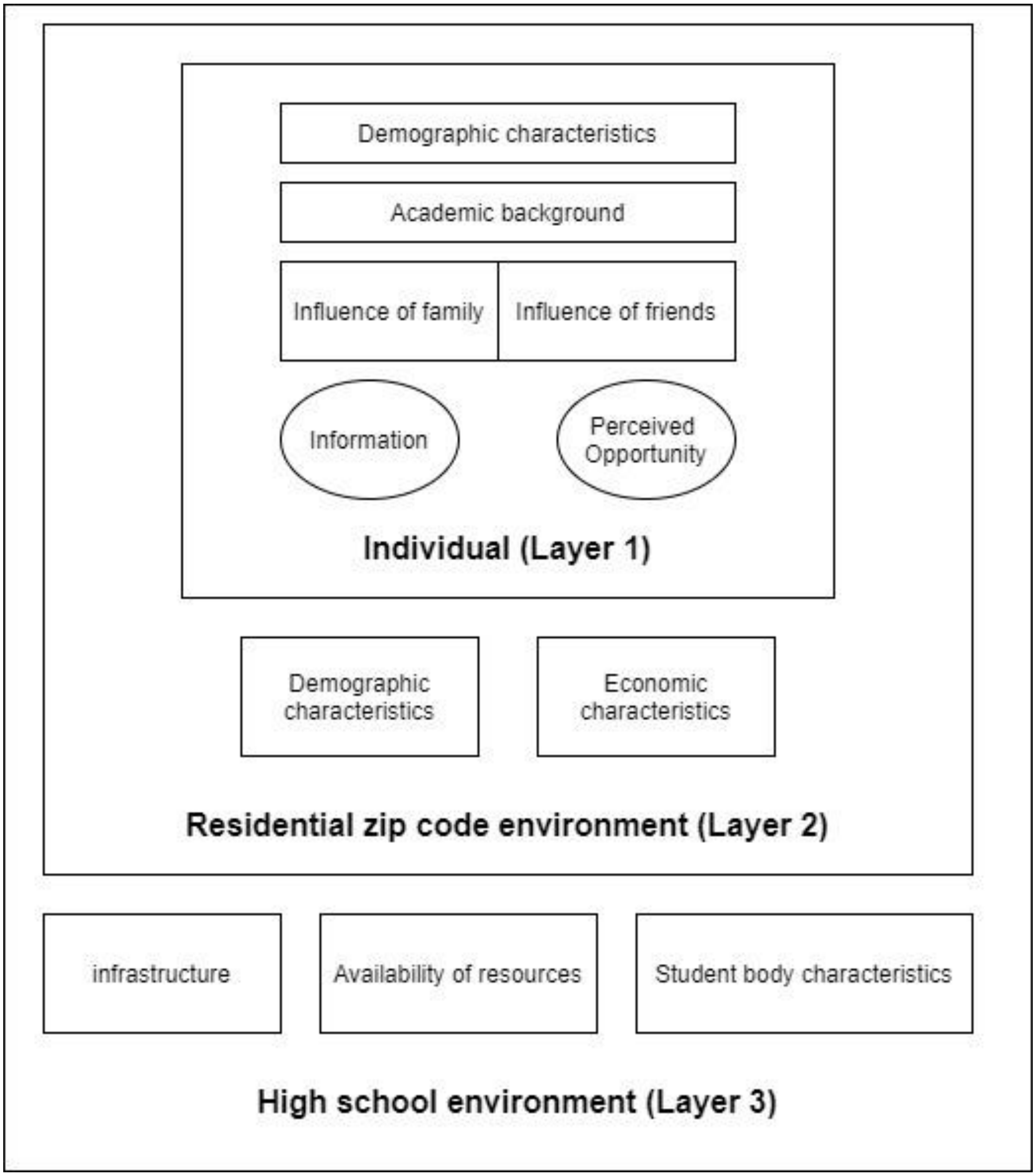

Figure 2.1 Adapted conceptual framework, mainly based on Perna (2006) framework, and integrating components from the Iloh (2018) college-going decision model.

Figure 2.2 Adapted Conceptual Framework for Research Question 3

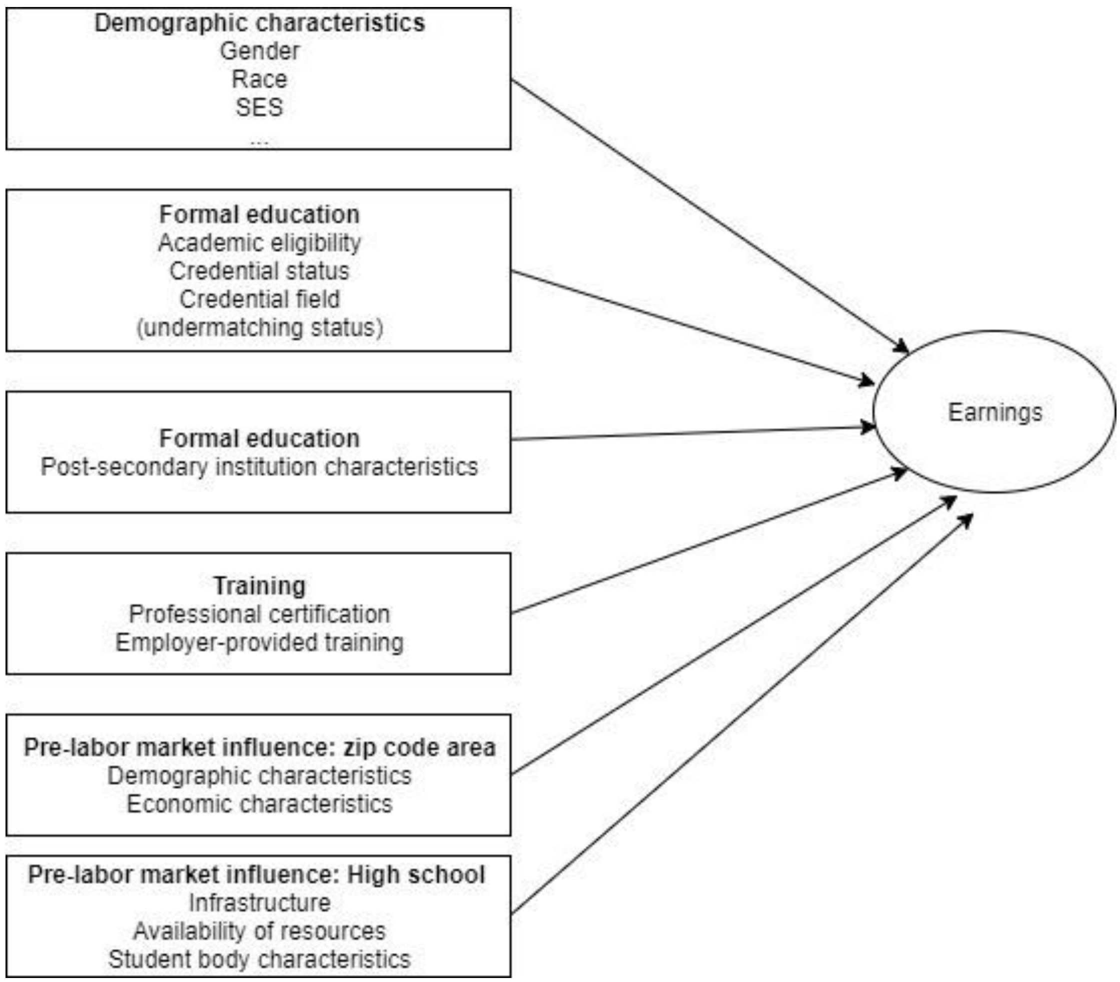

Figure 2.2 Adapted conceptual framework, based on human capital theory.

Chapter 3: Methodology

Research questions

The current study is guided by the following research questions:

1. How do students' background characteristics influence their chances of undermatching?

2. How do students' background characteristics, especially undermatching status, influence their probabilities of choosing STEM majors?

3. To what extent does undermatching influence students' labor market outcome (i.e. annual earnings from employment)?

    3.a. To what extent does choice of college major moderate the influence of undermatching on student labor market outcomes (i.e. annual earnings from employment)?

Research Hypothesis

1. To what extent do student characteristics influence their chances of undermatching?

*Hypothesis 1***:** I hypothesized that on individual level, racial background, parental SES, students' college choice priorities, and the number of college information sources would significantly influence students' chances of undermatching. I also hypothesized that zip code area level socioeconomic indicators would influence individual's undermatching status.

*Rationale 1***:** Researchers have found that socio-economic status is one of the most salient factors that are associated with being undermatched (Roderick et al., 2006; Allensworth, 2006; Roderick et al., 2008; Roderick et al., 2009; Roderick et al., 2011; Bowen et al., 2009; Smith et al., 2012; Smith et al., 2013; Freeman, 2016). Therefore I predict that lower socio-economic students have higher tendency to enroll in a less selective college, due to the reasons mentioned in literature review, such as limited college knowledge, inadequate information, insufficient planning, and lacking support to navigate the college application process (Hoxby & Avery, 2012; Hoxby & Turner, 2013; Bowen, et al., 2009; Rodrick et al., 2008; Rodrick et al., 2009). I also predicted that Black and Hispanic students are more likely to undermatch, as some studies have found that minority students, especially Black and Latino students, are significantly less likely to enroll in a matched school (Roderick, Coca, & Nagaoka, 2011; Bowen et al., 2009).

Students' college choice priorities, measured by questions asking the importance of certain factors when making choices about college, is also predicted to be significant. It is hypothesized that the more students value low-cost and vicinity, the more probable they will undermatch. Holland and Richard (1965) found that practical concerns, such as closeness to home and low cost, heavily influenced high school students' college choice, and this conclusion is also supported by other research (Bowers & Pugh, 1973). Therefore, I hypothesize that for undermatching students, these are particularly important factors that influenced their college choice.

This research also hypothesized that the more sources students have utilized in gaining college information, the less likely they will undermatch. College information sources often reflect the cultural capital that students possess. In addition to college counsellor in high school, the most commonly-used source, some students could also talk to parents about college plans, or have access to college publication websites, college search guides, or even college open-house visits. Additionally, the information component of the Iloh (2018) framework also suggests that students with access to multiple sources of credible information tend to make more informed decision about college-going. Therefore, it is hypothesized that the more ways students get access to college info, the more likely they will have a better understanding of college application, costs and finances, and less likely to undermatch.

I also hypothesize that zip code area characteristics would influence students' undermatching status. This is mainly based on Perna (2006)'s theoretical framework, that the economic and social context serves as external forces that influence student college decision. Even though currently there is no empirical evidence that suggests the influence of zip codes variables, a distantly related finding is that higher state college enrollment rates were related to state percentage of population that held a bachelor's degree or higher, state poverty rate, and the share of Hispanics in the polulation (St. John, 2004; Perna & Titus, 2004). Therefore, I predict that zip code area level characteristics would influence students' higher education decisions, thus influencing their undermatching status.

2. How do students' background characteristics, especially undermatching status, influence their probabilities of choosing STEM majors?

*Hypothesis 2***:** I hypothesized that after controlling for high school and zip code area characteristics, parental SES, and students' college choice priorities on job placement, would significantly influence students' chances of choosing a STEM major. I also hypothesized that undermatched students are less likely to choose STEM majors. And lastly, I hypothesized that zip code area socioeconomic indicators would influence individual's choice of college major.
*Rationale 2***:** Even though previous studies have found that underrepresented minority students, especially Hispanic and African American students, are less likely to graduate from STEM majors (Simpson, 2001; Berryman, 1983; Maple & Stage, 1991; Mullen, 2001; Powell, 1990; Museus, Palmer, Davis, & Maramba, 2011), Hispanic and African students report similar initial aspiration of choosing a STEM as their White and Asian peers (Herrera & Hurtado, 2011).

Therefore, I hypothesize that race will not significantly influence students' chances of choosing a STEM major.

On the other hand from previous literature, the influence of students' socioeconomic background has been found to be significant, even though there is no agreed conclusion about the mechanism of such influence. Early theories (Gamoran & Mare 1989; Hallinan, 1992) assume that students from higher SES families would choose college fields that predict better economic returns, as their parents would secure the best position in the education system for them. On the contrary, other research work (Kohn & Schooler, 1983; Davies & Guppy, 1997) found that students coming from lower SES background are more likely to treat higher education as a chance to move upward, while higher SES background students value rewards other than financial benefit. Therefore, I hypothesized that students coming from lower SES background are more likely to choose STEM majors, which are usually more lucrative than social science and humanities majors.

I also hypothesized that the more importance they place on postsecondary institutions' low expenses, availability of financial aid, and post gradation job placement, the more they would choose STEM majors. This is a hypothesis based on the "reputation" that STEM degree holders earn more (Beede, Julian, Khan, Lehrman, McKittrick, G., Langdon, D., & Doms, M. E. (2011), and as a result, students who are aspired to choose STEM for this reason might also be those who place importance on the financial aspects of postsecondary institutions.

The hypothesis of zip code area characteristics follows the same logic as the one in rationale for research question one. Meanwhile, as these aforementioned factors are also the ones that could potentially predict undermatching status, I also hypothesize that undermatched students are less likely to choose STEM majors. However, one thing to note is that, after controlling for these factors, being undermatched itself would not influence individuals' probabilities of choosing STEM majors.

3. To what extent does undermatching influence students' labor market outcome (i.e. annual earnings from employment)?

    3.a. To what extent does choice of college major moderate the influence of undermatching on student labor market outcomes (i.e. annual earnings from employment)?

*Hypothesis 3***:** I hypothesized that undermatching would negatively influence students' annual earnings from employment.

*Hypothesis 3.a***.** I hypothesized that majoring in STEM would alleviate the negative influence that undermatching status has on students' annual earnings from employment.

Rationale 3: Early studies showed that college selectivity has a significant positive effect on graduates' earnings (James, Alsalam, Conaty, & To, 1989; Fox, 1993; Brewer, Eide, & Ehrenberg, 1999; Monks, 2000). However, newer studies that used advanced techniques such as propensity score matching showed a weak or insignificant impact of college selectivity (Dale & Krueger, 2002; Black &Smith, 2004; Long, 2008). Even though currently there are no uniform conclusions regarding the effect of undermatching, some recent studies did demonstrate such effect (Hoekstra, 2009; Hughes, 2013) using different methodology (eg. regression discontinuity approach). Therefore, I hypothesize that undermatching will significantly decrease students' annual earnings.

*Rationale 3a***:** Major field of study has long been recognized as a factor that greatly influences college graduates' labor market outcomes (Rumberger & Thomas, 1993; Thomas 2003). More lucrative majors have a relatively specific and well-defined body of content knowledge, and focus on methods of inquiry requiring high levels of quantitative or scientific skills" (Wolniak, Seifert, Reed & Pascarella, 2008, p.125).

STEM graduates tend to have higher salaries than students who major in other disciplines (Pascarella &Terenizini, 2005; Carnevale, Cheah, & Strohl, 2012). More importantly, previous research has found that the earning effect of background characteristics such as gender, parents' education, and family income, are not uniformly moderated by education attainment across majors. For example, among Health Sciences and Social Sciences majors the moderating effect of level of educational attainment is significant, while among STEM majors the effect is inconsequential.

More importantly and directly related to hypothesis 3.a, research has found that college quality significantly influences the effect of academic major on earnings (Zhang & Thomas, 2005; Eide, Hilmer & Showalter, 2015). For example, business and social sciences graduates' earnings differ greatly by college quality; however for engineering majors, college quality does not significantly determine their earnings. Therefore, I hypothesize that majoring in STEM will alleviate the negative influence that undermatching status has on students' annual earnings from employment.

Data sources

ELS 2002

The individual level data sources for my dissertation is the Educational Longitudinal Study of 2002 (ELS 2002) data from the National Center for Education Statistics (NCES). The ELS 2002 is a nationally representative, longitudinal study of 10th graders in 2002, and followed through their secondary and postsecondary years. In addition to the 2002 base year survey, it has three follow-up student surveys (2004, 2006, and 2012 respectively), as well as high school transcript and college transcript (restricted use). This longitudinal dataset includes variables such as students' demographic backgrounds, family backgrounds, high school experiences, college application and choices, college experience, and labor market outcomes. The base year survey has a sample size of over 15,000 students from over 750 schools. The first follow-up survey was conducted in 2004, when most sample students were high school seniors while others were in other grades, dropped out, or completed high school early. Among students from the Base Year sample, 12,400 were still in the 750 schools included in the Base Year sample, 1,100 transferred to a different school, and 1,300 completed early or dropped out. The first follow-up transcript study collected high school transcripts for all students from their base year school, and these transcripts provide information on students' course completion, grades, attendance, SAT/ACT scores, and so forth from ninth grade to twelfth grade. The second follow-up survey was conducted in 2006, with all sample members who responded in the Base Year and/or the First Follow-up included. Many sampled students were in college up to the sophomore year, and many others were employed and may have never attended college. For high school dropouts many have earned a General Education Development (GED) or other equivalency certificate or be working on a GED. The latest follow-up occurred in 2012, 6 years after the Second Follow-up or 8 years after high school graduation. College transcripts were also obtained, including all courses completed for every college attended, and financial aid data for all years in college and every source.

ACS 2005

The zip code level data source for my dissertation is the American Community Survey (ACS) by the United States Census Bureau. The ACS is a longitudinal household survey that gathers information about ancestry, educational attainment, income, language proficiency, migration, disability, employment, housing characteristics, and more. Since 2005 is the year when the majority of ELS 2002 sample entered college, I use the 2005 data from ACS.

The 2005 ACS sampled approximately 2.9 million housing unit addresses annually stateside. The coverage rate is 98.5% for housing units, and 95.1% for total population; ACS coverage rate is calculated as the ratio of the ACS population or housing estimate of an area or group to the independent estimate for that area or group, times 100. The response rate is 97.3% nationwide.

IPEDS

The postsecondary institution characteristics data for my dissertation come from the Integrated Postsecondary Education Data System (IPEDS). IPEDS is a large-scale survey that collects institution-level data from postsecondary institutions in the United States by the NCES, and consists of twelve interrelated survey components. The 12 survey components include institutional characteristics, admissions, 12-month enrollment, completions, graduation rates, outcome measures, student financial aid, academic libraries, finance, and human resources.

The completion of all IPEDS surveys is mandatory for all institutions that participate in, or are applicants for participation in, any federal financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended. This mandatory participation requirement consequently results in nearly 100% response rate for each IPEDS survey component. In addition to the mandatory participants, the IPEDS database also includes institutions that do not participate in Title IV financial aid programs.

Analytic sample

For research question 1 and 2, factors influencing students' chances of undermatching, and chances of choosing a STEM major, the analytic sample is limited to students who were enrolled in post-secondary education (including four-year colleges and two-year colleges) two years after high school graduation, regardless of whether they earned a bachelor's degree from that institution. To account for unequal probabilities of selection and the problem of non-response, analytic weights provided by NCES will be utilized (Ingels et al., 2007). The analytic weight that used is F2BYWT, which is the "…second follow-up panel weight for all sample members who responded in the second follow-up and responded in the base year, or who were base-year nonrespondents but for whom the base-year classification variables were collected in the first follow-up and their base-year test scores imputed" (Ingels et al., 2007, p.152). The ultimate sample size is 9050.

Research question 3 examines to what extent does undermatching influence students' labor market outcome, and how does majoring in STEM moderate the influence of undermatching. As the labor market outcome data come from the third follow-up survey, the analytic weight used is F3F1PNLWT, which is the "Estimates based on third follow-up data in combination with first follow-up data (or second follow-up data) where the estimates are meant to represent students enrolled in 10th grade in the spring of 2002 or students enrolled in 12th grade in the spring of 2004" (Ingels et al., 2014, p.72). Furthermore, the analytic sample is limited to individuals who 1) have had post-secondary education, regardless of whether they had earned a degree, 2) were working full-time, at least 49 weeks and 30 hours per week, 3) reported annual income of $10000 or higher (an arbitrary criteria to eliminate outliers). The ultimate sample size is reduced to 3860.

Operationalize undermatching

Institutional selectivity

This study follows the practice of most previous undermatch research, and uses the collapsed Barron’s institutional competitive index as the definition of selectivity (Bowen, et al., 2009; Rodrick et al., 2008; Rodrick et al., 2009; Smith et al., 2013; Rodriguez, 2015). Barron’s index rates four-year postsecondary institutions based on their admission rate, GPA thresholds, and the standardized test scores of entering freshman class (Barron's Educational Series, 2006). The original Barron’s index is consisted of seven levels of selectivity, ranging from Non-competitive to Most Competitive. For the purpose of accommodating specific samples, some researchers utilized the collapsed Barron’s categories. Roderick et al.(2006; 2008; 2009) recoded from the Barron’s and added two-year institutions, producing a five-tier ratings (Two Year, Non-selective, Somewhat Selective, Selective, and Very Selective). Smith et al. (2012), Hughes (2013) and Fosnacht (2015) also follow this grouping method. Belasco and Trivette (2012) utilized similar method, but placed two-year institutions into the non-selective group, along with non-selective four-year colleges.

Previous researchers have identified the advantage of collapsing Barron’s categories on both ends: doing so improves statistical power (Belasco & Trivette, 2012). Beyond the statistical concern, there are “definitional advantages” (Belasco & Trivette, 2012, p.11). Both most competitive and highly competitive institutions are extremely selective, admitting less than half of the applicants, and have an average SAT scores among incoming student freshman that are adequately high. Therefore, it would make little sense to identify a student as undermatched, if they enroll in a Barron’s Highly Selective institution while they are eligible to attend a Barron‘s Most Competitive college. Moreover, the empirical research by Rodriguez (2015) shows that collapsing Barron’s categories has little influence on deciding the overall undermatch rates.

Similar reason support the decision to group the Barron's Less Competitive and Non-competitive four-year institutions into the non-selective category. Many less competitive four-year institutions practice open admission that resembles the practice of non-selective four-year institutions, and both categories of institutions admit students that are not eligible for selective four-year colleges (Belasco & Trivette, 2012). In summary, this study groups all institutions into four categories, including very selective, selective, somewhat selective, non-selective, and 2-year institutions.

Student qualifications

In current literature three methods have been used to decide students' academic qualifications: enrollment rate method, predicted probability method, and acceptance rate method (Rodriguez, 2015). The method this dissertation uses is the acceptance rate method. The enrollment method, as used in Roderick et al. (2006), utilized students' final enrollment information to decide their qualifications. This method involves grouping students by combination of SAT scores and GPA, and then the selectivity level that students of each group most attended is selected as the qualification level of that group. This method is suitable for researchers that do not have access to application data (Rodriguez, 2015). The predicted probability method, first utilized by Smith et al. (2012), takes a different approach. This method utilized logistic regressions to predict the probability of admission based on students' academic characteristics, such as honors-weighted GPA, ACT or SAT scores, and whether the student participates in Advanced Placement (AP) or International Baccalaureate (IB) coursework. However, the major shortcoming of this approach is that, a logistic regression model may determine that a student with a very low GPA but high SAT score have access to selective institutions, while in the real setting it is highly impossible. Many institutions carry out a holistic admission policy, and normally accept students who have shown adequate competency on every important criterion. Therefore the non-parametric method might avoid the risk of overestimating or underestimating the probability of students being accepted to institutions of specific selectivity level.

The third method is the acceptance rate method, and this is the approach this research is taking. As employed by Belasco and Trivette (2012), Roderick et al. (2009) and Bowen et al. (2009), this method defines an eligibility criteria based upon actual admission outcomes for institutions of each level of selectivity. Emphasizing on the combination of admission standards, this method is specifically carried out as below: For a particular combination of SAT (or converted ACT) score and high school GPA, if more than 90 percent of applicants were admitted into a specific level of selectivity, then all students who have the same or higher scores than the combination should get access to institutions of that selectivity level. Table 3.1 presents the example of very selective level, and among students who applied to at least one very selective institution and who have a combination a GPA between 3.5 and 3.79, and an SAT score between 1200 and 1300, 93.3% were admitted into at least one *very selective* institution, Therefore, all students with the same or higher combination of SAT (or ACT equivalent) and GPA were deemed to have access to a very selective institution.

Table 3.1. Determining access to selectivity: *very selective*

Very selective	GPA							
SAT	<2.0	2.0-2.29	2.3-2.59	2.6-2.89	2.9-3.19	3.2-3.49	3.5-3.79	3.8-4.1
≤800	13.2	22.2	48.6	17.2	*	*	*	*
801-900	*	26.5	25.0	40.5	44.4	43.3	*	*
901-1000	*	*	55.6	45.0	52.8	46.3	60.8	85.7
1001-1100	*	*	72.5	57.7	60.8	73.1	85.2	78.6
1101-1200	*	*	61.3	72.6	82.0	78.6	86.0	94.8
1201-1300	*	*	*	81.3	77.4	92.0	93.3	95.1
1301-1400	*	*	*	*	91.2	92.4	96.2	98.2
1401-1600	*	*	*	*	93.2	92.2	97.6	98.7

Variables

RQ1.Outcome Variable: Undermatching status

The first research question is, how are student background characteristics influencing their chances of undermatching? Therefore, the dependent variable is the college matching status after high school graduation. Independent variables are drawn, based on the conceptual framework in chapter two (Figure 2.1). According to Perna (2006) and Iloh (2018), the conceptual framework includes three layers, individual layers, residential zip code environment layer, and high school environment layer. Specifically, individual layer variables include demographic characteristics, academic performances, influence of family, and influence of friends, information, and perceived opportunity. Zip code area characteristics include several regional demographic and economic indicators, such as percentage of Bachelor's degree or higher, poverty rate, and average annual income. High school variables include high school characteristics, and student body composition. It should be noted that due to data limitation (not big enough sample size), the original high school and zip code cross-classified multilevel model was not feasible. Instead, a two-level student-high school level model was created, while zip code variables were treated as student level (level 1) variables. Detailed information regarding the revised model would be further explained below in the analysis technique section.

Demographics. Demographic variables include gender (BYSEX), race (BYRACE_R), SES (BYSES1QU), and language status (BYSTLANG). Race dummy variables were recoded from the NCES restricted-use data (BYRACE_R), which separates Native Hawaiians/Pacific Islanders from Asians. The original BYRACE_R variable includes American Indian/Alaska Native, Asian, Black or African American, Hispanic, more than one race, Native Hawaii/Pacific Islander, and White. This study did not include two categories due to small sample sizes, American Indian/Alaska Native (n=130), and Native Hawaii/Pacific Islander (n=61). The SES (BYSES1QU) variable is a composite variable based on five equally weighted, standardized components: SES is based on five equally weighted, standardized components: father's/guardian's education (FATHED), mother's/guardian's education (MOTHED), family income (INCOME), father's/guardian's occupation (OCCUFATH), and mother's/guardian's occupation (OCCUMOTH).

Academic background. Academic background variables include high school GPA (F1RGP), composite SAT or ACT equivalent score (TXEESATC), total AP/IB courses (F1MATHSE), math self-efficacy (F1MATHSE), and expected level of academic achievement (F1STEXP). Math self-efficacy (F1MATHSE) is a scale of the student's self-efficacy in math, created from five items, including "can do excellent job on math tests", "can understand difficult math texts", "can understand difficult math class", "can do excellent job on math assignments", and "can master math class skills". The coefficient of reliability for the scale is 0.91. Expected level of academic achievement (F1STEXP) includes 8 categories, with 1=less than high school graduation, and 8=PhD, MD, or other advanced degree.

College application. College application includes number of institutions applied to (F1S50), whether applied for financial aid (F2B04), post-sec school's low expenses important to respondent (F1S52A), availability of post-sec financial aid important to respondent (F1S52B), post-sec school's job placement record important to respondent (F1S52I), college information source-personal, and college information source-formal. College information source-personal was the sum of four items, "has gone to parent for college entrance information" (F1S48D), "has gone to sibling for college entrance information" (F1S48E), "has gone to other relative for college entrance information" (F1S48F),"has gone to friend for college entrance information"

(F1S48G). College information source-formal was the sum of seven items, including counselor (F1S48A) , college representatives (F1S48H), college publications/websites (F1S48I), college search guides (F1S48J), school library(F1S48K), public library(F1S48L), and college library(F1S4M8).

Family and friends' influence. Family influence includes two composite variables, parents provide advices on academics, and discussion with parents. Parents provide advices on academics is a composite variable constructed based on three items: Parents provide advice about selecting courses or programs (BYP56A), parents provide advice about plans for college entrance exams (BYP56B), and parents provide advice about applying to college/school after high school (BYP56C). The reliability for this 3-item scale is 0.724. Discussion with parents is a composite variable constructed based on five items: discuss school courses with parents (F1S64A), discuss things studied in class with parents (F1S64C), discussed grades with parents (F1S64D), discussed preparation for ACT/SAT with parents (F1S64G), discussed going to college with parents (F1S64H). The reliability for this 5-item scale is 0.776. Friends' influence include two variables, number of friends who consider grades very important (BYFRGRIM), and number of friends planning to attend 4-year college/university (F1S65D).

Zip code area influence. A total of five zip code demographic or economic indicators were selected. Educational attainment (S1501 from ACS), is the percentage of bachelor's degree or higher among population 25 years old and over. Language status (S1601 from ACS), is the percentage of speak English very well among population 5 years and over. Poverty rate (S1702 from ACS) is the percentage of families below poverty level. Income (S1901 from ACS) is mean households income. The last variable is percentage of White among the total population (DP1).

High school characteristics. High school characteristics include high school control (BYSCTRL), high school urbanicity (BYURBAN), total student enrollment (F1A01), student/teacher ratio (CP04STRO), percent of minority students (CP04PMIN), percent of student body is LEP or non-English proficient (F1A22B), and percent graduates went to 4-year colleges (F1A19A). Also, one aggregated school SES variable was created, by collapsing the student SES composite variable (BYSES1) by high school.

RQ2.Outcome Variable: Major choice

The second research question seeks to understand how do background factors, especially undermatching status, influence students' major choice. The dependent variable is the field of study students declared in 2006, the second year in college (F2MAJOR2). Due to the large number of majors identified in the ELS 2002 data, based upon the definition by Department of Homeland Security, the major choice variable was broadly divided to two categories: STEM, and non-STEM. STEM majors include Biological and biomedical sciences, Computer/info sciences/support technology, Engineering technologies/technicians, Mathematics and statistics, Physical sciences, Health professions/clinical sciences, and Science technologies/technicians. The full list of the STEM majors is included in the Appendix A. Independent variables are the same as research question 1, except that it also adds the dependent variable from research question one, the undermatching status.

RQ3.Outcome Variable: Employment earnings

Questions 3 and 3.a seek to understand the influence undermatching has on students' labor market outcomes, and the moderating effect of college major. The dependent variable is employment earnings, measured by the question, "about how much did you earn from employment in 2011 before taxes and all other deductions? Please include all wages, salaries, income from a business or farm, commissions, and tips you earned in 2011." As a side note, 2011 is 7 years after respondents graduated from high school, or 3 years after respondents graduated from a four-year higher education institution (if they did graduate).

Independent variables were selected based upon the conceptual framework shown in Figure 2.2, and it includes six components: demographic characteristics, pre-labor market influence-residential zip code area, pre-labor market influence-high school, formal education, postsecondary institution characteristics, and training.

Demographics. Similar to the demographics variables included in research question 1, research question 3 considers gender (BYSEX), race (BYRACE_R), SES (BYSES1QU), and language status (BYSTLANG), racial background (BYRACE_R), and SES quartile (BYSES1QU). In addition, since whether or not an individual is currently married has been found to be statistically associated with earnings (Cornwell & Rupert, 1997; Schoeni, 1995), research question 3 also added a dummy variable recoded from marital status (F3D01), indicating whether the respondent was married in 2012.

Zip code area influence. A total of five zip code demographic or economic indicators were selected. Educational attainment (S1501 from ACS), is the percentage of bachelor's degree or higher among population 25 years old and over. Language status (S1601 from ACS), is the percentage of speak English very well among population 5 years and over. Poverty rate (S1702 from ACS) is the percentage of families below poverty level. Income (S1901 from ACS) is mean households income. The last variable is percentage of White among the total population (DP1).

High school characteristics. High school characteristics variables include high school control (BYSCTRL), high school urbanicity (BYURBAN), total student enrollment (F1A01), student/teacher ratio (CP04STRO), percent of minority students (CP04PMIN), percent of student body is LEP or non-English proficient (F1A22B), and percent graduates went to 4-year colleges (F1A19A). Also, one aggregated school SES variable was created, by collapsing the student SES composite variable (BYSES1) by high school.

Formal education. The formal education component captures the credential status individuals hold. Four factors are included: credential field, academic eligibility level, credential level, and undermatching status, Credential field is a dummy variable recoded from the field of study of highest/only credential (F3ICREDGEN_1) as of 2012. Credential field is thus broadly divided into two categories: *obtaining a credential in STEM*, and *did not obtain a credential in STEM*. The *did not obtain a credential in STEM* includes the situation of obtaining a credential in non-STEM, and did not obtain a degree. STEM majors include Biological and biomedical sciences, Computer/info sciences/support technology, Engineering technologies/technicians, Mathematics and statistics, Physical sciences, Health professions/clinical sciences, and Science technologies/technicians, and the full list could be seen in Appendix A.

Academic eligibility level is decided using the aforementioned acceptance rate method. For a particular combination of SAT (or converted ACT) score and high school GPA, if more than 90 percent of applicants were admitted into a specific level of selectivity, then all students who have the same or higher scores than the combination should get access to institutions of that selectivity level. Four eligibility levels are thus created: *very selective*, *selective*, *somewhat selective*, and *nonselective and 2-year institutions*.

Credential level pertains to the level of credentials that student obtained. This variable is recoded from credential type of highest/only credential (F3ICREDTYPE_1), and indicated whether a student earned a bachelor's degree or higher, obtained an associate's degree or undergraduate certificate/diploma, or did not obtain a credential. And lastly, the undermatching status is the dependent variable from research question 1.

Postsecondary institution characteristics. To control for the influence of postsecondary institutions, several postsecondary institution characteristics indicators from the IPEDS are included. Total enrollment number (EFFY2005- fyrace24) is the 12-month unduplicated headcount. Percentage of White students is calculated by dividing White non-Hispanic count (EFFY2005- fyrace22) by the overall students count (EFFY2005- fyrace24). Similarly, Percentage of female students is calculated by dividing count of women (EFFY2005- fyrace16) by the overall student count (EFFY2005- fyrace24). I also included postsecondary tuition and required fees, which is the sum of in-state tuition (IC2005-AY-tuition2) and in-state required fees ((IC2005-AY-fee2). In addition, I included postsecondary institution controls (F3ISECTR), sectors (F3ISECTR), and HBCU (F3IHBCU) from the ELS 2002 third follow-up questionnaire.

Training. Training after formal education has been found to influence earnings as well, therefore two related variables are included. The first variable is certificate status (F3A27), indicating whether the respondent has a professional certification or license, either certified/licensed by state, professional organization, or industry/company/some other organization. The employer-provided training (F3B35) is another variable included here.

Analysis Techniques

Missing data. Firstly, cases with missing value for the outcome variables or demographic characteristics (i.e., gender and race) were deleted from the sample. In order to maintain as much participants in the analytic sample as possible, the expectation maximization (EM) algorithm was used to impute and replace missing values for all other continuous variables used in this study. The EM algorithm was used to find maximum likelihood (ML) estimates, shows more advantage than the less accurate missing values replacement such as mean replacement, and has been widely used in social science studies (Graham, Hofer, & MacKinnon, 1996; Little & Rubin, 1989). Missing data analysis shows that most of the variables had small percentage of missing data. The SAT composite or ACT equivalent score had 41% missing cases, and as the imputed values were estimated using other variables in the dataset, caution should be taken while interpreting results related to this measure.

RQ1.Background characteristics influencing undermatching

Research question 1 examines how are students' chances of undermatching are influenced by their background characteristics. To answer this question, descriptive analysis and Hierarchical Generalized Linear Model (HGLM) were conducted. Cross-tabulations with chi-squared tests were first used to compare the academic eligibility level across racial background and SES quartiles. Additionally, proportion tables were created to reflect the proportion of matching status (undermatched, matched, overmatched) across specific combinations of academic eligibility level and racial background/SES quartiles. Descriptive statistics were also shown to describe the final analytic sample used for multilevel modeling, presenting mean, standard deviation, minimum and maximum for each of the independent and dependent variables.

Exploratory factor analysis was conducted to reduce the number of independent variables used in the HGLM model. Principal axis factoring with promax rotation was used, which relaxes the constraint that the factors are uncorrelated with one another and therefore improves the fit to simple structure (Russell, 2002). To ensure internal reliability, the criteria for accepting a factor is that 1) within-factor variables must have loadings at .50 or higher, 2) all factors had an eigenvalue higher than 1.0, and 3) Cronbach's alpha of .70 or higher (Bland & Altman, 1997). The two factors created met the above criteria, including parents provided advice on academics, and discussed with parents about academics. Factor loadings, all scale measures, and cronhanch's alpha is presented in Appendix A2.

The second analysis method is HGLM, which suits nicely the need to understand what factors influence students' chances of undermatching in the context of high school and residential area. HGLM is based on Generalized Linear Model, which deals with observations that are not normally distributed, including binomial, multinomial, and Poisson. HGLM takes into account the hierarchical structure of data. In HGLM, samples are grouped into clusters; samples in the same cluster share similarities and are positively correlated.

With that being said, HGLM is appropriate for this study for both practical and methodological reasons. Firstly, the ELS 2002 study data utilized a two-stage sampling, with high schools selected first, then tenth-grade students selected randomly within each school. Data were collected for both individual students and high schools they came from; hierarchical linear modeling (HLM) recognizes the existence of such data hierarchies and allows for residual components at each level (Goldstein, Browne, & Rasbash, 2002). Moreover, the college-going decision model based on Perna (2006) and Iloh (2018) points to its "nesting" nature as well, with students nested within high school, and further within broader context.

Secondly, there are several methodological advantages of using HLM over traditional single-level regression, one being that HLM simultaneously considers variables at both student level and high school level to account for the clustering effect of students within schools. Therefore, the effects of both student-level and high-school level variables could be separately estimated. In addition, using traditional single level regression model on hierarchical data could lead to the standard error of regression coefficients being underestimated, thus increasing the possibility of type I error, or "false positive" (saying a parameter is significant while it actually is not; Goldstein, Browne, & Rasbash, 2002). Moreover, because HLM utilizes maximum likelihood techniques instead of ordinary least squares, which performs better when the sample is consisted of unequal groups (Hox & Van de Schoot, 2017; Raudenbush & Bryk, 2002). Therefore, HLM was the most appropriate technique for this study.

Using HLM technique also involves centering and weighting, in order to obtain more accurate and interpretable estimates. For this study, all continuous variables were centered at the grand means; therefore, the intercept would be the expected outcome for an "average" person in the population (Hox & Van de Schoot, 2017; Raudenbush & Bryk, 2002). All dichotomous variables were left uncentered. Additionally, models in the HLM 8 software were all weighted using corresponding panel weights.

For research question one, in order to understand the predictive power of student-level predictors and high school-level predictors, I conducted the analysis in blocks. To measure the between-school variance, an unconditional model, or in other words, a null model without any predictors was first conducted. The HGLM level-1 model is Bernouli and uses a logit function to predict the likelihood of undermatching for student *i* in school *j* (equation 1).

$$\eta_{ij} = \log[\phi_{ij}/(1 - \phi_{ij})] = \beta_{0j} \tag{1}$$

ϕ_{ij} is the probability of undermatching for student *i* in school *j*, and η_{ij} is the log odds, or likelihood of undermatching. Level 2 model is specified as:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00}) \tag{2}$$

The high school average on the outcome measure (undermachin), β_{0j}, is a function of the average log-odds of undermatching across all high schools, γ_{00}. u_{0j} is a random effect unique to each high school. The covariance estimates from the null model could then be used to calculate the intra-class correlation (ICC), which represents the proportion of variance between groups. In other words, this measures the extent to which students' average likelihood of undermatching varies across high schools. Even though the outcome variable here is dichotomous, and thus diminishes the accuracy and interpretability of the ICC, ICC is still produced as it still could help us understand the extent of variance between high schools. As with the single-level logistic regression model, the level 1 error variance is now heteroscedastic, the ICC for HGLM could be calculated by assuming that the level 1 error variance is $\pi^2/3$. Therefore, the ICC is estimated by the formula:

$$\text{ICC}= \tau_{00} /(\tau_{00} + \pi^2/3) \tag{3}$$

Starting from the null model, blocks of independent variables were added to the model one by one, following the adapted conceptual framework in Figure 2.1. The blocks were added in the following order: demographics, individual background (including academic background, college application, family influence, and friends influence), and residential zip code characteristics. Finally, level-2 variables were added to the model, to examine the effect of school-level predictors with the presence of student-level variables. The final HGLM student-level model can thus be represented by the following equation:

$$\begin{aligned}\log[\phi_{ij}/(1 - \phi_{ij})] &= \beta_{0j} + \beta_{1j}\,(\text{Demographics})_{ij} \\ &+ \beta_{2j}\,(\text{Academic background})_{ij} + \beta_{3j}\,(\text{College application})_{ij} \\ &+ \beta_{4j}\,(\text{Family influence})_{ij} + \beta_{5j}\,(\text{Friends influence})_{ij} \\ &+ \beta_{6j}\,(\text{Residential zip code})_{ij} \end{aligned} \tag{4}$$

And the high-school model is described by the following equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}{*}(\textit{High school characteristics})_{j} + u_{0j} \tag{5}$$

All level-2 parameters are constrained to be fixed across schools, with $\beta_{\rho j} = \gamma_{\rho 0}$, and $\rho = 1$ to n student level variables.

RQ2.Factors influencing students' major choice

Research question 2 seeks to understand the extent to which students' background factors, especially undermatching status influenced students' major choice. Descriptive analysis were first conducted. Proportion tables were created to reflect the proportion of choosing STEM across specific combinations of undermatching status and racial background/SES quartiles. The HGLM model is almost the same as presented in equation (4) and (5), except that the dependent variable became the dichotomous variable of whether chose a STEM major or not, and undermatching status was entered into the model first before adding other variables. The final HGLM student-level model can thus be represented by the following equation:

$$\log[\phi_{ij}/(1 - \phi_{ij})] = \beta_{0j} + \beta_{1j}\,(\text{undermatching status})_{ij}$$
$$+ \beta_{2j}\,(\text{Demographics})_{ij} + \beta_{3j}\,(\text{Academic background})_{ij}$$
$$+ \beta_{4j}\,(\text{College application})_{ij} + \beta_{5j}\,(\text{Family influence})_{ij}$$
$$+ \beta_{6j}\,(\text{Friends influence})_{ij} + \beta_{7j}\,(\text{Residential zip code})_{ij} \tag{6}$$

And the high-school model is described by the following equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\textit{High school characteristics})_{j} + u_{0j} \tag{7}$$

All level-2 parameters are constrained to be fixed across schools, with $\beta_{\rho j} = \gamma_{\rho 0}$, and ρ = 1 to n student level variables.

*RQ*3.Effect of undermatching on employment earnings

Research question 3 examines the influence of undermatching on students' labor market outcomes, and specifically how does majoring in STEM moderate the influence of undermatching. The original intended analytic method for research question three is propensity score weighting. However, balance could not be achieved, even after reducing the variable set considered for matching to include only gender, race, SES and GPA variables (see Appendix C for the balance check table). Shadish, Clark and Steiner (2008) suggest that propensity score models with limited covariates might not perform as well as a more traditional model with observable characteristics as covariates. And the analysis by Eagan et al. (2013) further demonstrates that more standard estimation techniques (such as HLM), if the models include a rich set of covariates that extend beyond demographics, perform just as well as propensity score models. Therefore, I switched the analysis method to HLM.

Descriptive analysis was first conducted to compare the sample used in research question 3 and the previous sample used in research question 1 and 2, including mean, standard deviation, and range. Descriptive analysis also examines how individual's employment earnings vary by undermatching status and background factors (eg. race, gender, SES, academic background, credential status), along with F-test to test significant differences in the average earnings across undermatching status in the population.

Because the dependent variable is annual earnings from employment (continuous), Hierarchical Linear Modeling (HLM) was used. Similar to research question 1 and 2, a null model was run first, in order to determine the ICC. The formula for ICC in HLM model is calculated using the equation:

$$\text{ICC} = \tau_{00} / (\sigma^2 + \tau_{00}) \tag{8}$$

τ_{00} is the variance at level-2, and σ^2 is the variance at level-1.

Next, blocks of variables were entered one by one into the model, in the following order: core terms (including undermatching status, credential field, and interaction), individual background (demographics, high school characteristics, zip code area characteristics, and training), formal education (credential level and academic eligibility), and postsecondary institution characteristics. Individual annual earnings variable (the dependent variable) was log-transformed to adjust for skewness and improve model specification (Montgomery, Peck, & Vining, 2012). Zip code annual average household income, and postsecondary tuition and fees were also log-transformed. The final HLM model is represented in the following equation:

$$\begin{aligned} \text{LogIncome} = {} & \beta_{0j} + \beta_{1j}\,(\text{core terms})_{ij} \\ & + \beta_{2j}\,(\text{Demographics})_{ij} + \beta_{3j}\,(\text{Residential zip code})_{ij} \\ & + \beta_{4j}\,(\text{Training})_{ij} + \beta_{5j}\,(\text{Formal education})_{ij} \\ & + \beta_{6j}\,(\text{Postsecondary institution})_{ij} + r_{ij} \end{aligned} \tag{9}$$

And the high-school model is described by the following equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\textit{High school characteristics})_j + u_{0j} \tag{10}$$

$\beta_{\rho j} = \gamma_{\rho 0}$, and ρ = 1 to n student level variables.

Limitation

While this study offers a comprehensive examination of factors that predict undermatching and choosing STEM, and how they interact to predict earnings, it is important to acknowledge the limitations. Ideally a cross-classified HGLM model would be more suitable to understand the influence of both high schools and zip code areas, and it fits well with the conceptual model of college-going decisions that based upon Perna (2006) and Iloh (2018). Students in the sample come not only from different high schools, but also from different communities; not all students from the same school come from the same community, and not all students from the same community go to the same school. Schools are not nested in communities; rather, the relationship is cross-classified. However, the reality of ELS 2002 dataset makes using a cross-classified model impossible, as the average sample size over zip code is too small (around 2). To still account for zip code characteristics influence, I included zip code characteristics variables as student level variables, and created a student-school two level model instead.

Simply fitting a two-level model to cross-classified data could result in misattributing response variation to the included levels (van Landeghem et al., 2005; Moerbeek, 2004; van den Noortgate et al., 2005; Tranmer and Steele, 2001). In other words, including neighborhood level variables in student-school two-level model but ignoring neighborhood as a level in the model could lead to an underestimation of standard errors on these neighborhood variables, thus increasing the chance of making type I error (Leckie, 2013). Still, for this study including zip code variables could add to our understanding of the undermatching issue, even though conclusion regarding the zip code characteristics variables should be treated with caution. Future studies, when possible, could select a few zip code areas and further examine how residential zip code interact with high schools in influencing student outcomes. The measurement of zip code area characteristics could also include more aspects. The Galster and Killen (1995)'s geography of opportunity model suggests that the way geography influences individuals are through education, legal labor market, the criminal justice system, the illegal labor market, and social welfare. The current study could only capture a few characteristics of the first two aspect-education and labor market, and future studies could consider adding data from more resources to measure, for example, crime rate, illegal labor market, and unemployment rate differences between individuals of different educational attainment.

Secondly, the idea of undermatch as the underpinning for the whole research is not without problem. While it is widely accepted that higher education system is highly stratified, no consensus has been reached as to the measurement of the stratification of collegiate institutions (Lucas, 2001). There is no doubt that the most selective institutions could be easily identified, but among less selective colleges, institutional ranking are more fluid and indefinite, especially for those local and regional institutions (Bastedo & Bowman, 2010; Bastedo & Flaster, 2014). Even though this study follows numerous previous research (Bowen, et al., 2009; Rodrick et al., 2008; Rodrick et al., 2009; Smith et al., 2013; Rodriguez, 2015) and utilizes the Barron's admissions competitiveness index, the problem of fluidity among less selective institutions still exists. Consequently, results that are related to lower selectivity institutions (in this study, somewhat selective and nonselective institutions) should be treated with caution. On the other hand, the presumptive eligibility in this study is determined solely by the combination of GPA and SAT/ACT scores. Real-life college admission usually implements holistic review, and considers factors more than these two academic indicators. As a result, the notion that "researchers could accurately predict who should be accepted" is problematic.

The operationalization of undermatching in this study is another limitation. This study only focuses on the match between students' presumptive eligibility and the selectivity level of first institution they attended. However, literature (Adelman, 1999) found that a large proportion of students did not stay in the same institution throughout their postsecondary journey, and there is increasing trend of new patterns of student attendance, including regular one-way transfer (complete part of the coursework at the first institution then transfer to finish at the second institution) , multiple attendance (back-and-forth enrollment among two or more institutions) and "double-dipping" (concurrent attendance at two institutions; de los Santos and Wright, 1990; Gose, 1995; McCormick, 2003). More importantly, the inclusion of two-year colleges as the fifth rung in higher education ladder also might cause problem. One important function of two-year colleges is transfer (Orfield & Paul, 1992), and it offers low-income, first-generation, and underrepresented minority students the low-cost alternative of first half of college education.

Therefore, future research should decompose the college path starting from initial institution, to graduating institution, and explore how undermatch is influencing in each part of the higher education experience.

Another limitation especially related to labor market outcomes is that the employment earnings reported in this study is only early-career earnings. The year 2011 is three year after most students in the 2002 cohort graduated from four-year colleges (if they did graduate). Even though research has indicated some connections between early-career earnings and mid-career earnings (Agnello & Hunt Jr, 1976), there is still variation in estimation, especially for lifelong earnings (Haider & Solon, 2006). Therefore, future studies should also breakdown the influence of undermatching on earnings at other time point of life.

Also, since this study utilizes pre-existing data, some variables used did not fully capture the construct that was originally intended to measure. For example, prior research (Baum & Schwartz, 2015) and the Iloh (2018) model suggested that the source of credible college information matters in students' college-going decisions. However, the items in the ELS 2002 questionnaire only capture whether or not a student has gone to certain individuals for college entrance information, leaving out measures about the quality of such interaction, and not to mention it is impossible to decide what is considered as "credible". Another example is parental involvement. The two scales utilized in this study are parents providing advices on academics, and students' discussion with parents about academics. Again, these two scales could at most measure the subjective judgment about frequency of discussion (never, sometimes, often), overlooking the quality of such discussion, and the actual frequency. Future studies on students' interaction with all parties should look to include qualitative measures of behaviors, or conduct qualitative studies to further understand the way and mechanism of specific behaviors.

Lastly, as the sample size of research question 3 is limited to those working full-time 3 years after college graduation (or 7 years after high school graduation), interpretation based on the results should be applied to the full-time working population. Since students of some STEM majors are more likely to pursue graduate education (e.g., biology and physics, Carnevale, Strohl & Melton, 2011), it is also possible that students from these majors are underrepresented in the sample. Moreover, there are a series of related problems this study could not answer, for example, to what extent does undermatching influence persistence and how does the influence differ by major, as well as the reason why undermatched student in STEM earn more in their early career. Future study should continue to explore the aforementioned problems.

CHAPTER 4: RESULTS

This chapter presents findings from descriptive analysis and hierarchical modeling techniques for each research question. The reporting of numbers follow the NCES guidelines, and are all rounded to the nearest tens place.

Examining How Well Institutional Selectivity Matches with Students' Academic Qualifications

The first question this study addressed related to the individual, high school, and residential (or neighborhood) factors associated with whether students undermatched when deciding where to enroll for their postsecondary education. Given the dichotomous nature of the dependent variable (undermatch) and the nested nature of the data, hierarchical generalized linear modeling represented the most appropriate statistical method to address this question.

Table 4.1 Description of Student-level and High School Level Variables in Analytic Sample (weighted n= 9050 students)

Individual level predictors	Min	Max	Mean	SD
Demographics				
Female	0.00	1.00	0.54	0.50
Asian(ref: White)	0.00	1.00	0.10	0.30
Black(ref: White)	0.00	1.00	0.11	0.32
Hispanic (ref: White)	0.00	1.00	0.11	0.31
Multiracial (ref: White)	0.00	1.00	0.05	0.21
Lowest quartile SES (ref: Highest quartile)	0.00	1.00	0.12	0.33
2nd lowest quartile SES (ref: Highest quartile)	0.00	1.00	0.22	0.41
3rd quartile SES (ref: Highest quartile)	0.00	1.00	0.28	0.45
First Language English	0.00	1.00	0.86	0.34
Academic background				
High school GPA	0.36	4.20	2.97	0.65
Composite SAT or ACT equivalent score(100)	4.20	16.00	9.95	2.06
Total AP/IB courses	0.00	18.00	1.07	1.88
Math self-efficacy	-2.04	1.85	0.05	0.91
Expected level of academic achievement	1.00	8.00	6.53	1.28

College application				
Number of institutions applied to	0.00	4.13	2.63	0.94
Post-sec school’s low expenses important to respondent	1.00	3.00	2.13	0.67
Availability of post-sec financial aid important to respondent	1.00	3.00	2.40	0.71
Post-sec school’s job placement record important to respondent	1.00	3.00	2.50	0.62
Whether applied for financial aid	0.00	1.00	0.74	0.43
College info Personal	0.00	4.00	1.79	1.10
College info Formal	0.00	7.00	2.16	1.12
Family influence				
Parents provide advice on academics	-1.81	1.29	0.10	0.88
Discussion with parents: academics	-2.70	1.61	0.16	0.84
Friends' influence				
# friends plan to attend 4-year college/university	1.00	5.00	3.64	0.94
# friends who consider grades very important	0.00	3.00	1.45	1.01
Residential zip code influence				
Zip % Bachelor's degree or higher (10%)	0.00	10.00	2.85	1.61
Zip % Speaks English well (10%)	3.26	10.00	9.25	0.96
Zip % Poverty (10%)	0.00	9.17	2.77	0.76
Zip % White (10%)	0.06	10.00	7.69	2.30
Zip Annual household income ($10000)	2.22	30.64	7.46	3.00
High School level predictors				
High School characteristics				
School control (Ref: public)	0.00	1.00	0.23	0.42
School urbanicity (Ref: urban)	0.00	1.00	0.67	0.47
% minority	0.00	100.00	31.55	31.67
Student/teacher ratio	1.52	54.17	16.78	5.11
Total student enrollment	0.00	4533	1235.6	822.07
% of graduates went to 4-year colleges	1.00	6.00	4.48	1.16
% of student body is LEP or non-English proficient	0.00	50.00	5.71	9.47
School aggregated SES	-0.81	1.4	0.12	0.38

Table 4.1 presents the descriptive statistics for all individual level and high school level variables in the analytic sample, including minimum, maximum, mean, and standard deviation. This sample comes from a larger sample of high school students, but only those who were still enrolled in a college or university the second year after high school graduation is included. In this sample, a majority of students in the sample identified as female (54%) and nearly three out

of five students (63%) identified their race as White. More than one third (38%) are from the highest quartile SES, and the vast majority of students reported English as their first language (86%). The average GPA reported is 2.97, and the average composite math/verbal SAT scores of participants fell just below 1,000 (995), while the average number of AP/IB courses taken is 1. Meanwhile, most participants expected to obtain a bachelor's degree (value of 6). Students in the sample on average applied to two to three institutions, and 74% of them applied for financial aid. The average number of personal college info source is close to two (1.79), while on average students have around two (2.16) formal sources of college info. The average zip code annual household income is $74600, and the bachelor's degree or higher percentage is 28.5%. As to the high schools that students come from, a majority is public (77%) and urban schools (67%). The average percentage of minority student is 31.5%. Percent of graduates went to 4-year colleges is categorical, and the common percentage is between 25% and 49%.

As a refresher, students' academic eligibility level was decided using the combination of their SAT (or converted ACT) score and high school GPA. If more than 90 percent of the applicants with a particular combination were accepted in a particular selectivity level, then all students who have the same or higher scores than the combination should be academically eligible for institutions of that selectivity level. Meanwhile, these categories are mutually exclusive, and each student case only contributes to the calculation once. For example, a student eligible to enroll in *very selective* institutions but choose to attend a *nonselective* college is treated as undermatched only in the *very selective* column, but not included in the *selective*, *somewhat selective* column. Detailed information of the 90% threshold SAT and GPA combination of each selectivity level is listed in table 4.2.

Table 4.2 Determining Academic Eligibility Level

Very selective				GPA				
SAT	<2.0	2.0-2.29	2.3-2.59	2.6-2.89	2.9-3.19	3.2-3.49	3.5-3.79	3.8-4.1
≤800	13.2	22.2	48.6	17.2	*	*	*	*
801-900	*	26.5	25.0	40.5	44.4	43.3	*	*
901-1000	*	*	55.6	45.0	52.8	46.3	60.8	85.7
1001-1100	*	*	72.5	57.7	60.8	73.1	85.2	78.6
1101-1200	*	*	61.3	72.6	82.0	78.6	86.0	94.8
1201-1300	*	*	*	81.3	77.4	92.0	93.3	95.1
1301-1400	*	*	*	*	91.2	92.4	96.2	98.2
1401-1600	*	*	*	*	93.2	92.2	97.6	98.7
Selective								
≤800	40.0	31.4	32.1	47.0	35.7	*	*	*
801-900	*	34.1	43.3	62.5	60.0	58.7	78.4	*
901-1000	*	44.3	63.8	77.3	80.2	85.4	91.5	90.5
1001-1100	*	73.9	68.4	86.0	86.5	92.3	94.9	97.7
1101-1200	*	*	79.6	90.9	95.6	93.9	96.2	100.0
1201-1300	*	*	*	91.1	91.2	94.2	97.4	99.6
1301-1400	*	*	*	*	95.6	93.7	99.2	100.0
1401-1600	*	*	*	*	*	*	98.3	100.0
Somewhat selective								
≤800	45.2	57.4	58.5	77.3	76.7	73.6	*	*
801-900	58.5	69.7	75.8	82.7	87.5	96.1	92.2	90.5
901-1000	75.5	84.8	86.8	91.7	91.0	94.5	94.1	97.5
1001-1100	77.3	87.8	90.9	93.9	97.0	98.0	98.4	97.9
1101-1200	*	*	96.7	93.4	97.1	98.6	96.6	99.1
1201-1300	*	*	*	98.3	100.0	98.0	98.1	98.3
1301-1400	*	*	*	*	*	97.6	90.4	100.0
1401-1600	*	*	*	*	*	*	*	100.0
Nonselective								
≤800	71.8	79.8	77.9	77.9	80.7	93.2	*	*
801-900	83.3	72.2	84.7	87.8	90.1	90.9	88.6	*
901-1000	*	89.3	94.8	91.1	88.7	90.2	92.9	*
1001-1100	*	96.0	91.4	94.6	95.5	94.3	94.9	97.5
1101-1200	*	*	*	86.7	88.9	88.0	94.5	91.1
1201-1300	*	*	*	*	*	*	81.3	98.2
1301-1400	*	*	*	*	*	*	*	95.7
1401-1600	*	*	*	*	*	*	*	*

Note1: Each cell shows the percentage of students with the specified combination of high school GPA and SAT scores (or ACT equivalent) who were admitted into the level of institution.
Note2: Cells with less than 20 cases are marked with asterisk (*).
Note3: Cells with percentage higher than 90% are marked gray. Less selective level are marked with a less grey scale.
Note4: The following cell, (GPA 2.9-3.19 & SAT 901-1000), (GPA 2.9-3.19 & SAT 1101-1200), (GPA 3.2-3.49 & SAT 1101-1200), (GPA 3.5-3.79 & SAT 801-900), and (GPA 3.5-3.79 & SAT 1201-1300), have admission rate less than 90% at Nonselective. But since students with such combination could be admitted into even higher selectivity institutions, they are considered to be able to be admitted by the Nonselective institution levels.
Source: ELS 2002

Before analyzing the factors that influence undermatch, it is necessary to first examine how students' academic eligibilities distributed across levels of institutional selectivity. As displayed in table 4.3, the overall percentage of students whose academic backgrounds suggest they are eligible to be admitted to and enroll in very selective institutions is 14.4%. Approximately one-fifth (21.2%) of students are qualified for selective institutions with roughly the same proportion (22.0%) demonstrating eligibility for admission to institutions classified as somewhat selective. Relatively few students (7.2%) had qualifications best suited for admission to nonselective institutions. This is due to the fact that, according to the aforementioned operationalization of students' academic eligibility level, the SAT and GPA combination at the 90% threshold in the somewhat selective level is only slightly different from that in the nonselective level (see table 4.2). More than one third (35.2%) of students only qualify for 2-year/open admission colleges.

Table 4.3 Cross tabulation of Academic Eligibility Level, by Race, for Weighted National Sample (n=9050)

	Very selective (n=1280) %	Selective (n=1960) %	Somewhat selective (n=1970) %	Nonselective (n=680) %	2-year (n=3160) %
Asian (n= 410)	24.1	23.7	19.5	6.2	26.6
Black (n=1170)	1.7	7.1	14.5	6.8	70.0
Hispanic (n=1160)	4.5	9.9	17.8	8.1	59.7
Multi (n=370)	10.2	21.1	22.3	8.7	37.6
White (n=5940)	18.1	26.4	24.7	7.8	23.0
Overall (n=9050)	14.4	21.2	22.0	7.2	35.2

* Pearson chi2(16) = 1.4e+03 Pr = 0.000

Table 4.3 shows detailed academically eligible level breakdown for each racial group. About a quarter of all Asian students have academic qualifications making them eligible for admission to the most selective institutions (24.1%), a figure that eclipses the proportion for White students by six points and that national average by 10 points. Nearly as many Asian students were eligible to enroll at institutions classified as *selective* (23.7%) compared to 26.4%

of White students. By contrast, Black and Latino students show different patterns regarding the selectivity level of colleges and universities that are best matched with their academic qualifications. Less than 5% of Black (1.7%) and Latino (4.5%) students had academic backgrounds that would have made them eligible for admission to *very selective* institutions. Instead, 70.0% of Black students and 59.7% of Latino students had qualifications best matched for enrollment at *two-year* institutions. In short, Asian and White students have a higher percentage that are academically eligible for *very selective* and *selective* institution, while Black and Hispanic students' academic qualifications are more concentrated in lower selectivity institutions.

Table 4.4. Matching status for Weighted National Sample, by Racial Background (n=9050)

	Overall	Very selective	Selective	Somewhat selective	Nonselective	2-year
Asian (n= 410)						
Proportion eligible	100	24.1	23.7	19.5	6.2	26.6
Match	52.5	63.8	35.1	34.3	18	79.9
Undermatch	30.1	36.2	42.1	41.4	44.1	0
Overmatch	17.4	0	22.8	24.3	37.9	20.1
Black (n=1170)						
Proportion eligible	100	1.7	7.1	14.5	6.8	70.0
Match	53.3	25.2	25.7	41.3	30.8	61.2
Undermatch	12.4	74.8	61.5	39.5	21.0	0
Overmatch	34.3	0	12.8	19.1	48.2	38.8
Hispanic (n=1160)						
Proportion eligible	100	4.5	9.9	17.8	8.1	59.7
Match	56.0	43.2	20.3	22.2	26.3	77.9
Undermatch	25.1	56.8	63.9	63.0	56.3	0
Overmatch	18.8	0	15.8	14.8	17.4	22.1
Multi (n=370)						
Proportion eligible	100	10.2	21.1	22.3	8.7	37.6
Match	42.3	37.9	22.0	34.4	30.2	61.7
Undermatch	30.7	62.1	56.4	45.3	26.8	0
Overmatch	26.9	0	21.2	20.3	43.0	38.3
White (n=5940)						
Proportion eligible	100	18.1	26.4	24.7	7.8	23.0
Match	41.9	35.8	27.3	35.9	15.7	77.5
Undermatch	44.4	64.2	61.2	51.4	58.3	0
Overmatch	13.7	0	11.4	12.7	26.0	22.1

Total (n=9050)						
Proportion eligible	100	14.4	21.2	22.0	7.2	35.2
Match	45.5	38.4	27.1	34.7	19.8	72.9
Undermatch	36.9	61.6	60.2	51.0	51.3	0
Overmatch	17.6	0	12.7	14.3	28.9	27.1

The rates of match, undermatch and overmatch by race/ethnicity for the weighted national sample are displayed in table 4.4. The proportion eligible row is copied from table 4.3, so that readers could see the over/undermatch rates vis-a-vis the proportion of particular racial group who actually are qualified or eligible for particular selectivity institution enrollment. One thing to note is that, different from table 4.3, the percentages shown here are not row percentages. Instead, the percentages are proportion of matched, undermatched and overmatched students with specified combination of racial background and academically eligible level. Within each racial group and within each column, the match, undermatch, and overmatch rate will add up to 100%. As a caution, because of low cell sizes (less than 30, highlighted), the interpretation should only be applied to the sample itself, instead of the national high school student population.

The overall national undermatch rate is 36.9%. This overall undermatch rate includes the group *two-year* institutions (which by definition has an undermatch rate of 0); therefore, even though at the first four selectivity levels, a majority of student undermatched, the overall undermatch rate is still far below 50% for the entire sample of students who entered a higher education institution within two years of graduating from high school. There is a positive relationship between academically eligible level and undermathing rate: higher academically eligible levels suggest those students have higher undermatching rate. Among students who are academically eligible for *very selective* institutions, a relatively large proportion undermatched (61.6%), and nearly the same proportion of students with academic eligibility for admission to *selective* institutions ultimately undermatched (60.2%). A slight majority of students with academic eligibility for enrollment at *somewhat selective* (51.0%) and *nonselective* institutions (51.3%) also undermatched. Students whose academic eligibility aligns with enrollment at two-year institutions cannot, by default, undermatch. In nature, this relationship is an artifact of the definition of undermatch—students that are academically eligible for higher levels simply have more levels to be undermatched.

Moreover, the undermatching rates by race and selectivity demonstrate more interesting results. The overall undermatching rate for Asian students is moderately lower (30.1%) than the national sample (36.9%), and Asian student that are academically eligible for *very selective* and *selective* institutions undermatch at a rate of 36.2% and 42.1% respectively, the lowest among all racial groups. Black students, on the other hand, show a completely different undermatching pattern. At first glance, Black students seem to undermatch at the lowest rate (12.4%), but a further analysis shed more light on this issue. Among Black students with academic eligibility for enrollment at *very selective* institutions, an exceptionally high percentage (74.8%, but be cautious of the small cell sample size) undermatched. Roughly three out of five (61.5%) Black students eligible to enroll at *selective* institutions ultimately decided to pursue degrees at less selective campuses. Still, because a large proportion of Black students were academically eligible for *2-year* institutions (see table 4.3, which has an undermaching rate of 0), the overall undermatch rate for Black students is extremely low.

Interestingly, the overall undermatching rate of White students is the highest among all five racial groups (44.4%). The undermatching rates for White students are similar to the national rate, but because of the large sample size of White students, and high percentage of White students that are academically eligible for *very selective* institutions, White students' overall undermatching rate becomes the highest.

Although many students undermatch with respect to their academic eligibility and ultimate decision regarding where to pursue a college degree, a number of others overmatch by gaining admission to and enrolling at institutions that, on paper, seem to be more selective than would be expected given their academic eligibility. Table 4.4 also shows rates of overmatch by race/ethnicity. As mentioned in chapter two, overmatch does not mean that students are under qualified; some institutions implement holistic review that best meet institutional goals, such as athletic ability, artistic talent, race, gender, legacy status, economically disadvantaged background, and personal experience (Espenshade, Chung, & Walling, 2004; Fetter, 1997). Therefore, the concept of overmatch here is solely based on academic background, and should not be interpreted as if related to enrollment qualification.

The general pattern is exactly the opposite of undermatching: students with higher levels of academic eligibility tend to be less likely to overmatch. This is also determined by the nature of data structure, similar to the mechanism of undermatching. The national overall overmatch rate is 17.6%; in other words, among all students who have enrolled in post-secondary institutions, 17.6% enrolled in institutions whose selectivity level is higher than their academic eligibility level. Thus, at the *very selective* level, by default no one overmatches. At the *selective* level and *somewhat selective* level, 12.7% and 14.3% overmatch respectively. The overmatch rate increases significantly at the *nonselective* and *2-year college* level – 28.9% overmatch at the *nonselective* level, and 27.1% overmatch at the *2-year* level. This means that among students who are only academically eligible for *nonselective* or *2-year colleges*, more than one-quarter of them were actually enrolled in a more selective institution.

The racial breakdown gives further information regarding the matching issue. Noticeably Black students have the highest overmatch rate (34.3%), much higher than the national average

(17.6%). But a careful examination of the overmatch pattern over different academic eligibility level shows a different insight. Overmatch rates among Black students academically eligible for *selective* and *somewhat selective* institutions track closely with the respective national averages; however, nearly half (48.2%) of Black students with qualifications best suited for *nonselective* institutions and more than a third (38.8%) with qualifications for *two-year* institutions ultimately overmatched when making their final choice of where to enroll. Since Black students are overrepresented in *nonselective* and *two-year* institutions, this drastically elevated their overall overmatch rate.

For Asian students overall, their overmatch rate is similar to the national average, but the pattern across levels of academic eligibility shows some slight variation relative to the national sample. Among Asian students who are academically eligible for *selective* institutions, 22.8% overmatched, meaning that they got into the *very selective* level institutions. This is much higher compared to the national overmatch rate at the *selective* level (12.7%). Multiracial students overmatch at the second highest rate (26.7%), but the pattern is quite unique. Different from Black students who have a high overall overmatch rate but low overmatch rate at *selective* level, multiracial students have the second highest overmatch rate (21.2%) not only at the *selective* level, but also all other levels. And finally, White students overmatch at the lowest rate (13.7%) across all five racial groups, and their overmatch rate is also the lowest at the *selective* level (11.4%) and *somewhat selective* level (12.7%).

Table 4.5 Cross tabulation of SES and Academically Eligible Level, for Weighted National Sample (n=9050)

	Very selective (n=1280) %	Selective (n=1960) %	Somewhat selective (n=1970) %	Nonselective (n=680) %	2-year (n=3160) %
Lowest (n= 1560)	3.7	12.0	16.3	10.1	57.8
2nd (n=2060)	7.4	16.1	21.7	8.1	46.7
3rd (n=2550)	12.4	22.5	25.4	7.7	32.0

Highest (n=2880)	26.1	29.5	22.6	5.1	16.7
Overall (n=9050)	14.4	21.2	22.0	7.2	35.2

* Pearson chi2(12) = 1.5e+03 Pr = 0.000

In addition to racial or ethnic background, socioeconomic status also stratifies the extent to which students overmatch or undermatch their choice of college with their academic eligibility. Table 4.5 is the cross tabulation of socioeconomic status and academically eligible level. As mentioned earlier, in the national sample, 14.4% are academically eligible for *very selective* institutions, 21.2% to *selective*, 23.0% to *somewhat selective*, 7.2% to *nonselective*, and 35.2% to *2-year* colleges. However, the patterns for students coming from different SES background are quite different.

It comes as no surprise that the most affluent students tend to have academic backgrounds that make them eligible for admission to more selective institutions. A majority of students from the highest SES quartile have academic backgrounds that make them eligible for enrollment at either *very selective* (26.1%) or *selective* (29.5%) institutions. By contrast, just one-third of students in the third SES quartile were eligible for *very selective* (12.4%) or *selective* (22.5%) institutions, and the numbers plummeted further for those in the second (7.4% and 16.1%, respectively) and bottom (3.7% and 12.0%, respectively) quartiles. By contrast, the majority of students coming from the bottom SES quartile are best matched for enrollment in *two-year* institutions (57.8%).

In general, there is clearly a relationship between academically eligible rate and SES background. High SES students have higher academically eligible rates at more selective institution, low SES students have higher academically eligible rates at less selective institutions. Further analysis, combined with the information of undermatch rate, will be presented in later sections.

Table 4.6 Matching status for Weighted National Sample, by Socioeconomic Background (n=9050)

	Overall	Very selective	Selective	Somewhat selective	Nonselective	2-year
Lowest (n= 1560)						
Proportion eligible	100	3.7	12.0	16.3	10.1	57.8
Match	55.8	34.3	23.2	27.4	22.4	78.4
Undermatch	26.5	65.7	70.3	58.7	54.8	0
Overmatch	17.7	0	6.5	13.9	22.8	21.6
2nd (n=2060)						
Proportion eligible	100	7.4	16.1	21.7	8.1	46.7
Match	47.7	25.3	19.2	28.3	18.3	73.4
Undermatch	35.3	74.7	73.9	61.3	60.1	0
Overmatch	17.0	0	6.9	10.4	21.6	26.6
3rd (n=2550)						
Proportion eligible	100	12.4	22.5	25.4	7.7	32.0
Match	42.4	24.6	26.0	36.2	22.3	70.2
Undermatch	41.0	75.4	64.3	52.1	51.7	0
Overmatch	16.6	0	9.7	11.7	26.0	29.8
Highest (n=2880)						
Proportion eligible	100	26.1	29.5	22.6	5.1	16.7
Match	42.1	46.6	31.6	39.9	15.2	64.6
Undermatch	39.6	53.4	49.9	40.0	38.2	0
Overmatch	18.2	0	18.5	20.1	46.6	35.5
Overall (n=9050)						
Proportion eligible	100	14.4	21.2	22.0	7.2	35.2
Match	45.5	38.4	27.1	34.7	19.8	72.9
Undermatch	36.9	61.6	60.2	51.0	51.3	0
Overmatch	17.6	0	12.7	14.3	28.9	27.1

Table 4.6 presents the matching status by socioeconomic background. The proportion eligible row is copied from table 4.5, so that readers could see the over/undermatch rates side by side with the proportion of particular SES group who actually are qualified or eligible for particular selectivity institution enrollment. From the undermatch rows, the overall national undermatch rate is 36.9%, and the rates are 61.6%, 60.2%, 51.0%, and 51.3% for *very selective, selective, somewhat selective*, and *nonselective* levels, as mentioned in previous section about race.

Although students in the bottom SES quartile are the least likely to have academic eligibility for enrollment in *very selective* institutions, nearly two-thirds of these students undermatched (65.7%). By contrast, nearly three-quarters of students in the second lowest SES quartile (74.7%) who are eligible to enroll in *very selective* institutions decide to pursue their degrees in less selective institutions, which is second highest among the quartiles. Students from the second lowest quartile undermatch at the highest rates for each of the other academic eligibility levels. More than three-quarters of students from the second highest SES quartile eligible for enrollment at *very selective* institutions ultimately undermatch (75.4%). The third quartile students, like those in the top quartile, are well-represented in the top three selectivity levels of eligibility, but these students may not have the financial resources to afford attending the most selective campuses, or cultural capital to understand how financial aid could help offset the costs of enrolling at more selective campuses.

On the other hand, the overall national overmatch rate is 17.6%, and it appears that this rate hardly varies across SES quartiles in the aggregate; however, variation emerges when looking at overmatch by both SES and academic eligibility as shown in Table 4.6. Just 6.5% of students from the poorest families who are academically eligible to enroll at *selective* institutions overmatch, the lowest of the four SES groups and less than half of the national rate of 12.7%. Across each level of academic eligibility, students from the bottom two SES quartiles are less likely than their more affluent peers to overmatch.

Notably, students from the highest SES quartile show a completely different pattern. Judging from the overall overmatch rate, there isn’t a big difference between the highest SES group (18.2%) and the national rate (17.6%). However, at every selectivity level, highest SES students overmatch at a rate much higher than the national rate. At the *selective* level, the overmatch rate for the highest SES students is 18.5%, much higher than the other SES groups and the national rate (12.7%). At *somewhat selective* level, the rate is 20.1% vs. national rate of 14.3%; at *nonselective* level, it's 46.6% vs. national rate 28.9%; then lastly at the *2-year* level, its 35.5% vs. the national 27.1%.

Factors That Influences Undermatching Status

Four sets of variables were entered into the model, and the nested-model results are presented in table 4.7. The first model is demographic only model, including gender, racial background, SES status, and language status. The second model adds other individual level variables, including academic background, college application, family and friends influence. The third model adds residential zip code characteristics, depicting the environment of the residential residential zip code that students come from. The final model includes high school characteristics, including school control, urbanicity, and characteristics of high school student body. Delta-p statistics are reported for statistically significant predictors in each model, which represent the probability change of undermatching as a result of a one-unit change in the predictor variable holding all other variables constant at their mean values. The calculation of delta-p statics follows the recommendation by Peterson (1985) and Cruce (2009).

A null model without any predictors was run first, to determine the alternative ICC. In the null model, the variance component is 0.383, which resulted in an alternative ICC of .104. This means that an estimated 10.4% of the variance in college undermatch can be attributed to high-school level variables at level 2. The variance component was also found to be significant, with χ^2 (746) = 1668.64, p <.001. Moreover, by comparing the variance component of the third model (residential zip code model)(0.169), and the final model that contains school-level variables

Table 4.7 Step-wise HGLM Model Predicting Undermatch (n = 9050 students, 750 High Schools)

Model	Demographics				Individual				Residential zip code				High school			
	Coef.	S.E	Sig.	Δ-p	Coef.	S.E	Sig.	Δ-p	Coef.	S.E	Sig.	Δ-p	Coef.	S.E	Sig.	Δ-p
Demographics																
Female	0.136	0.056	*	3.2	-0.142	0.071	*	-3.3	-0.127	0.071			-0.117	0.072		
Asian(ref: White)	-0.412	0.119	***	-9.0	-0.433	0.140	**	-9.4	-0.440	0.146	**	-9.5	-0.493	0.148	***	-10.6
Black(ref: White)	-1.582	0.116	***	-26.1	-0.522	0.135	***	-11.1	-0.583	0.148	***	-12.3	-0.604	0.152	***	-12.7
Hispanic (ref: White)	-0.697	0.111	***	-14.3	-0.153	0.131			-0.221	0.139			-0.246	0.142		
Multiracial (ref: White)	-0.546	0.144	***	-11.6	-0.232	0.169			-0.251	0.171			-0.265	0.173		
Lowest quartile SES (ref: Highest quartile)	-0.225	0.099	*	-5.1	0.273	0.124	*	6.5	0.186	0.126			0.078	0.130		
2nd lowest quartile SES (ref: Highest quartile)	-0.066	0.076			0.404	0.098	***	9.8	0.329	0.099	***	8.0	0.253	0.102	*	6.1
3rd quartile SES (ref: Highest quartile)	0.086	0.070			0.385	0.085	***	9.3	0.347	0.086	***	8.4	0.304	0.087	***	7.3
First Language English	0.259	0.114	*	6.2	0.008	0.134			0.025	0.137			0.035	0.138		
Academic Background																
High school GPA					2.015	0.086	***	44.6	1.963	0.087	***	43.8	1.917	0.088	***	43.1
Composite SAT or ACT equivalent score(100)					0.436	0.029	***	10.6	0.461	0.030	***	11.2	0.494	0.031	***	12.0
Total AP/IB courses					-0.278	0.023	***	-6.2	-0.278	0.024	***	-6.2	-0.289	0.024	***	-6.4
Math self-efficacy					-0.119	0.039	**	-2.7	-0.121	0.039	**	-2.8	-0.119	0.039	**	-2.7
Expected level of academic achievement					-0.065	0.029	*	-1.5	-0.063	0.029	*	-1.5	-0.066	0.029	*	-1.5
College application																
Number of institutions applied to					-0.446	0.043	***	-9.7	-0.431	0.044	***	-9.4	-0.417	0.044	***	-9.1
Post-sec school's low expenses important to respondent					0.234	0.057	***	5.6	0.239	0.058	***	5.7	0.242	0.058	***	5.8
Availability of post-sec financial aid important to					0.151	0.059	**	3.4	0.127	0.059	*	3.0	0.123	0.059	*	2.9
Post-sec school's job placement record important to					-0.174	0.054	**	-4.0	-0.168	0.054	**	-3.8	-0.164	0.055	**	-3.7
Whether applied for financial aid					-0.144	0.086			-0.185	0.087	*	-4.2	-0.203	0.088	*	-4.6
College info Personal					-0.014	0.031			-0.008	0.032			-0.002	0.032		
College info Formal					0.041	0.034			0.043	0.034			0.042	0.034		
Family influence																
Parents provide advices on academics					-0.044	0.039			-0.044	0.039			-0.042	0.039		
Discussion with parents: academics					0.072	0.043			0.073	0.043			0.074	0.044		

Friends influence																
# friends plan to attend 4-year college/university					-0.184	0.041	***	-4.2	-0.170	0.041	***	-3.9	-0.134	0.042	**	-3.1
# friends who consider grades very important					-0.034	0.034			-0.038	0.034			-0.043	0.034		
Residential zip code characteristics																
zip % Bachelor's degree or higher (10%)									-0.128	0.055	*	-3.0	-0.062	0.056		
zip % Speaks English well (10%)									-0.055	0.049			-0.032	0.056		
zip % Poverty (10%)									-0.016	0.055			-0.018	0.055		
zip % White (10%)									-0.005	0.023			0.000	0.028		
zip Annual income ($10000)									-0.052	0.016	***	-1.2	-0.034	0.017	*	-0.8
High school characteristics																
School control (Ref: public)													-0.100	0.128		
School urbanicity (Ref: urban)													-0.024	0.089		
% minority													0.001	0.002		
Student/teacher ratio													0.030	0.010	**	0.7
Total student enrollment													0.000	0.000		
% of graduates went to 4-year colleges													-0.105	0.045	*	-2.4
% of student body is LEP or non-English proficient													0.000	0.006		
School aggregated SES													-0.404	0.163	*	-8.8
Intercept	-0.583	0.033	***		-0.788	0.038	***		-0.796	0.037	***		-0.782	0.038	***	
Variance Component(S.D)	0.249	0.499			0.214	0.463			0.169	0.411			0.135	0.367		
Reliability	0.354				0.262				0.222				0.187			
-2 Log Likelihood	25362.66				24687.44				24806.92				24871.62			

**p*<.05, ***p*<.01, ****p*<.001

(0.135), it could be found that the high school characteristics explain an additional 20.1% of the between-school variance in rates of undermatching variance.

Demographic factors. In the demographic only model, gender was originally a significant predictor of undermatching, and female students are 3.2% more likely to undermatch. However, after controlling for other individual level predictors, females are 3.3% less likely than male students to undermatch. Upon further examination, it appears that the introduction of high school GPA to the model resulted in the sign reversal associated with gender. In other words, although descriptive statistics might show that women umdermatch at higher rates than men, that fact is in part due to the fact that women earn significantly higher grades than men. After accounting for high school grades, women have a significantly lower likelihood of undermatching relative to men who have similar academic backgrounds in high school. Gender loses its significance by the third model when residential zip code characteristics are added to the model, and it appears that the percentage of residents in the zip code area with a bachelor's degree or higher and the average income of the zip code accounted for the elimination of gender as a significant predictor of undermatch, as women in this sample tended to come from communities with higher incomes and higher levels of education.

In terms of racial background, when only demographics factors enter the model, Asian, Black, Hispanic, and Multiracial students are all less likely than White students to undermatch, which follows the findings presented earlier in this chapter. After other individual background factors were added, Asian and Black students remain less likely to undermatch, and undermatch rates among Hispanic and Multiracial students no longer significantly differ from White students. Part of the reason Hispanic and multiracial students undermatched at lower rates than White students was due to the fact that they also had lower grades and test scores compared to White students; once the model accounted for those characteristics, Hispanic and multiracial students no longer significantly differed from their White peers. In the final model, Asian students' probability of undermatching is 10.6 percentage points lower than White students' probability, and Black students' probability is 12.7 percentage points lower than their White peers. This is also in accordance with Smith, Pender, and Howell (2013) that Black and Asian students are less likely than White students to undermatch,; however they also found Hispanic students are marginally less likely to undermatch.

Now examining the SES differences in terms of undermatching, students from the lowest quartile are 5.1% less likely than those from the highest quartile to undermatch. However, an interesting sign reversal happens after other individual background factors are entered in the model- students from the lowest quartile SES are 6.5% more likely than students from the highest SES quartile to undermatch. In other words, students from the lowest quartile SES are initially less likely to undermatch, primarily due to their concentration in the lower levels of academic eligibility. After controlling for individual background factors, namely high school grades and standardized test scores, students in the bottom SES quartile undermatch at higher rates than their similarly capable but more affluent peers. In other words, although it appears from a univariate perspective that students from the lowest SES quartile undermatch at lower rates than their more affluent peers in the top SES quartile, this statistic is in large part due to the fact that less affluent students have lower grades and test scores than students in the top quartile, which by default means they qualify to enroll in less selective institutions thereby having fewer opportunities to undermatch. Once the model considers these differences, we see that the poorest students in this sample undermatch at significantly higher rates than the most affluent students in the sample. This difference becomes nonsignificant after accounting for residential zip code characteristics, as students from less affluent families tend to come from communities with lower levels of education and lower income levels, both of which were also significant predictors of the decision to undermatch.

Students from the second lowest SES quartile and second highest SES quartile show a different pattern. When there are only demographic factors in the model, these students are not significantly different from those coming from the highest SES background in terms of the probability to undermatch. However, after individual background factors enter the model, students second lowest SES quartile and second highest quartile students are more likely than highest quartile SES students to undermatch, and the effects remain significant even after residential zip code factors and highs school factors are included in the model. This change to the respective coefficients occurs in large part once high school GPA, the number of AP/IB courses, and standardized test scores are controlled. Thus, simple comparisons of undermatch rates between students in the second and third quartiles with those in the highest SES quartile would suggest no significant difference primarily because of the lower grades, number of courses, and scores associated with students in the lower SES quartiles. Once these differences in academic eligibility are controlled, the model suggests these students in the second and third SES quartiles are significantly more likely to undermatch relative to the most affluent students in the sample.

The last demographic factor, language status, first emerged as a marginally significant predictor. When other demographic factors are controlled for, students whose first language is English are 6.2 percentage points more likely than those whose first language is not English to undermatch. This might look like a surprising result, given common belief and previous findings (Drake, 2014; Fry, 2007; Rampey, Dion, & Donahue, 2009) that language minority status generally has a negative impact on student achievement. This indeed is the case, and it is actually part of the reason why language minority students are less likely to undermatch. Students for whom English is not their first language tend to have lower GPA (mean of 2.90 compared to 2.98 of English as first language student) and SAT scores (mean of 954 compared to 1000 of English as first language student); therefore, students whose first language is not English are less likely to be eligible to attend the most selective institutions making, by default, their probability to undermatch lower. This explanation could also be validated by the individual background model. When individual background factors are also controlled for, there is no significant difference between students of different language status, in terms of the probability to undermatch.

Academic background. All the academic background factors are significant, even after residential zip code and high school variables are also controlled. High school GPA and SAT scores are all positively related to undermatch, which means that the higher a student's GPA and SAT score, the more likely they would undermatch. This finding directly connects to the definition of undermatch—students who have higher GPA and SAT scores by definition are academically eligible for admission to and enrollment at more selective institutions; thus, they have more levels to enroll to be undermatched.

On the other hand, AP/IB courses, math self-efficacy, and expected level of academic achievement, are all negatively related to undermatch. In other words, students taking more AP/IB classes, possessing higher match self-efficacy, or expecting to have higher levels of academic achievement, are less likely to undermatch. However, note that a simple correlation between these variables and undermatch reveals a positive relationship, meaning that there was a sign signal due to multicollinearity with GPA and/or SAT test scores.

Remarkably, the sign reversal of taking AP/IB classes is particularly interesting. The simple correlation shows a positive relationship between taking AP/IB classes and undermatch ($r=0.09$, $p<.001$), while in the HGLM model the coefficient is negative, showing that students who take more AP/IB courses undermatch less often than we would expect them to given their high grades and test scores. This finding potentially suggests that either critical information about applying to college is shared and/or learned when students take more of these AP/IB courses, or simply taking such courses serves as an indicator that students may have some additional forms of cultural capital that better inform their final selection of which college to attend.

College application. A few college application factors are also found be significantly associated with whether students undermatched when selecting which college to attend. Firstly, students who applied to more institutions had a significantly lower probability of undermatching. Specifically, every additional college application a student submitted corresponded with an estimated 9.1 percentage point reduction in the probability of undermatching. This result is understandable, as submitting more applications likely increases the number of acceptances a student may receive, and more acceptances likely provide students with a larger pool of campuses to determine which would make the best match with their backgrounds, skillsets, and personal preferences.

In addition to the number of applications, the findings from the model indicate that students who place a high value on the cost of attendance and availability of financial aid, it is more likely that they will undermatch. In particular, this finding likely explains why students from less affluent families undermatch at higher rates after controlling for all variables in the model, as these students are more sensitive to price concerns and more dependent on the extent and nature of financial aid. This is similar to Belasco and Trivitte (2015)'s finding, and students from the low SES background might be deterred by the sticker price, and worry about finance more. As a result, they might just enroll in colleges with lower sticker price (usually lower selectivity insitutions), and compromise for financial aid.

By contrast, students who gave job placement greater consideration when selecting their college tended to be less likely to undermatch. This is interesting, and it is consistent with the rising trend that when choosing colleges, students are emphasizing college graduates' employment prospects (Eagan, Stolzenberg, Ramirez, Aragon, Suchard, & Rios-Aguilar, 2016), and in general, graduates from more selective institutions tend to have higher earnings (Rumberger & Thomas, 1993; Drydakis, 2016).

Students who applied for financial aid were significantly less likely to undermatch, compared to those who did not apply for financial aid, but this effect is significant only once residential zip code characteristics and high school contexts are included in the model. The explanation lies in the correlation among undermatching, zip code income, and financial aid application. The correlation between zip code income and financial aid application is negative ($r = -0.17$), meaning that if students come from a high income zip code, they are less likely to apply for financial aid. On the other hand, applying for financial aid is positively related to undermatching ($r = 0.079$), and zip code income is negatively related to undermatching ($r = -0.07$). Therefore, although undermatch rates for those who applied for financial aid and those who did not apply for aid might not differ when comparing them directly, significant differences emerge once the characteristics of students' communities are controlled in the model. Because more students who apply for financial aid tend to come from lower-income communities, the effect of financial aid is masked until the model accounts for the fact that coming from a more affluent community corresponds with a reduced likelihood of undermatching, as students from more affluent communities likely have better access to information and resources related to the benefits of attending a more prestigious institution. The fact that, after account for differences in undermatching based on the mean income of students' home zip code, financial aid applicants are expected to have a significantly lower likelihood of undermatching suggests that the mere act of applying for financial aid may signal that financial aid applicants have some level of cultural capital or college-going knowledge that may distinguish them from their peers who do not apply for aid.

College information sources, however, are not found to impact the probability to undermatch, including both personal sources and formal sources. The number of personal sources, namely those personal-based relationships, include parents, siblings, other relatives, and friends. Moreover, the number of formal college info sources, including high school counselors, college representatives, college publications/websites, college search guides, and school or public libraries, do not seem to influence whether students undermatch. Considering that previous literature has confirmed the importance of college information sources, specifically parents' and siblings' understanding of college application (Ceja, 2006; Kimura-Walsh, Yamamura, Griffin, & Allen, 2009), as well as formal sources (McKillip, Rawls & Barry, 2012; King, 1996; Cabrera, & La Nasa, 2001), one explanation of the contradicting result is that the influences of personal and formal sources are wrapped up in other individual background factors, such as the financial aid and AP/IB course taking. Another reason might be that the quantity of sources does not matter. Rather, it is possible that the quality of college information sources is a more reliable predictor than simply the quantity of sources, but this dataset did not include any proxy measures of information quality.

Family & Friends Influence. Neither family influence variable significantly predicted undermatch. While prior research has revealed that parental involvement could benefit academic experience through improving academic achievement, school attendance, parent/student interaction, grades and aspiration (Greenwood & Hickman, 1991; Jeynes, 2007; Fann, McClafferty Jarsky, & McDonough, 2009), one possible explanation is that parental involvement indirectly, rather than directly, relates to undermatch. Parental involvement tends to be associated with stronger academic qualifications, which can increase a student's likelihood of undermatching, but this study's findings suggest no direct influence of parents on whether students choose a college well-matched with their academic eligibility.

In terms of friends' influence, on average every one more friend that plan to attend 4-year college/university would decrease the probability of undermatching by 3.1 percentage points. However, the number of friends who consider grades very to be important did not significantly relate to whether students undermatched. It is possible that students who consider grades to be very important are also peers who plan to go to college, and thus multicollinearirty between these two measures may have contributed to the lack of significance with respect to having friends who value good grades.

Residential zip code characteristics. Among all residential zip code characteristics, the percentage of residents who have earned at least a bachelor's degree and the overall average income of students' communities represented the only variables to significantly predict whether students undermatched when selecting their college. Students from communities more densely populated by college graduates tend to undermatch at lower rates. The same pattern applies to the average income. However, after high school characteristics are controlled for, the educational attainment of communities becomes a marginally significant predictor, and the average income no longer significantly relates to undermatching. The reduction in the predictive power of contextual measures related to students' zip codes after the introduction of high school characteristic to the model suggests that more proximal contexts have greater saliency in predicting students' behavioral patterns than more distal (i.e., residential communities or neighborhoods) influences.

High school characteristics. Only three variables related to students' high school contexts significantly predicted students' probability to undermatch. Firstly, attending high schools with higher student-teacher ratios, which would typically translate into larger class sizes and less individual attention to students, correspond to higher probabilities of undermatching. For every one unit increase in the student/teacher ratio translates to a 0.7 percentage point increase in the probability of undermatching. High student teacher ratio means that each student is getting less attention from teachers, which may lead to less informed decision making, even among some of the most academically talented students.

Students who attended high schools with higher rates of graduates enrolling in four-year colleges tended to have a lower likelihood of undermatching. In essence, having a stronger college-going culture in high school encourages students to attend four-year institutions and likely shapes students' academic aspirations, thus reducing undermatching rate. Lastly, the average SES of students' high schools correlated with reduced odds of undermatching. Having access to more resources within high school contexts appears to school aggregated SES is also negatively related to undermatch, and the higher school aggregated SES is, the less likely students are to undermatch. Therefore, in addition to the effect of individual SES, the SES of other students in the same high school also influences the chances of undermatch.

Factors That Influences Major Choices (STEM vs. Non-STEM)

Having established the characteristics associated with whether students undermatch when selecting where to enroll in college, this study now considers whether undermatching correlates with whether students decide to major in one of the STEM disciplines. This section presents the descriptive and hierarchical generalized linear model analysis on factors that influence students' major choice (STEM vs. not choosing STEM), including student level variables, high school level variables, and residential zip code level variables.

Table 4.8 shows that in general, only a small percentage of students choose STEM majors (21.1%), and a majority (78.9%) either did not choose a STEM major or had not declared a major by spring of 2006. As a side note, this research did not exclude those who have not declared a major in 2006, because it is possible that those who undermatched may be more "adrift" or out of place, thus making them potentially less likely to select a major. Therefore, to better examine the relationship between undermatch and selecting a STEM major, the population in this sample is the same as that in research question one, which includes students who were still enrolled in college in 2006 (ELS2002 cohort's second year in college).

Table 4.8 Proportion of Students Majoring in STEM, by Alignment between Institutional Selectivity and Academic Eligibility and Race/Ethnicity, and SES quartile (n=9050)

	Overall	Matched	Undermatched	Overmatched	P-value	Sig.
By Race/Ethnicity						
Asian (n= 410)	32.4	29.8	36.9	32.7	0.301	
Black (n=1170)	22.9	18.7	29.5	27.4	0.023	*
Hispanic (n=1160)	15.5	13.9	17.0	19.3	0.059	
Multi (n=370)	20.8	18.2	25.9	18.1	0.330	
White (n=5940)	21.1	18.0	25.1	15.9	0.000	***
Overall (n=9050)	21.2	18.1	25.4	19.9	0.000	***
By SES						
Lowest (n= 1560)	21.4	17.3	28.0	23.7	0.009	**

2nd (n=2060)	19.9	16.4	22.8	21.6	0.037	*
3rd (n=2550)	20.1	16.9	24.8	17.7	0.000	***
Highest (n=2880)	22.6	20.8	26.0	19.2	0.000	***
Overall (n=9050)	21.2	18.1	25.4	19.9	0.000	***

Moreover, it seems that there is a moderate difference in the proportion of students majoring in who matched, undermatched, and overmatched. Undermatched students tend to major in STEM disciplines (25.4%) at higher rates than their peers where the selectivity of their first institution either matched (18.1%) or exceeded (19.9%) their estimated academic eligibility for admission. Chi-square tests examine the relationship between *whether chose a STEM major* and *whether undermatched*, and the result displays the overall STEM rate differences across matching status in each racial group. No significant differences are apparent across the status of matched, undermatched, and overmatched for Asian, Hispanic, and Multiracial students; but for Black and White students, the differences are significant, showing that the proportion of students choosing STEM major significantly differs among matched, undermatched and overmatched population. In terms of SES quartile, for all SES quartile groups the STEM rate differences across matching status are all significant.

Considerable variation in deciding to major in STEM by race/ethnicity is also clearly shown in Table 4.8. Nearly one-third of Among Asian students (32.4%) choose STEM majors, the highest among all five racial groups. By contrast, just 15.5% of Hispanic students decided to pursue a degree in a STEM discipline. About one in four (22.9%) Black students chose to major in STEM, and about one out of five Multiracial and White students majored in STEM (20.8% and 21.1% respectively). On the other hand, no clear pattern exists between SES quartile and the rate at which students decide to major in STEM. Overall, the proportion of student who chose STEM major is not significantly different among four SES quartiles. Still, one caveat is that Chi-square test is extremely sensitive to sample sizes; therefore, while the chi-square test results provide some insights, further HGLM analysis sheds more light on the issue.

Hierarchical Generalized Linear Model Results

To better understand how different sets of variables are influencing students' probability to choose a STEM major, and to further explore the relationship between undermatching and STEM, a hierarchical generalized linear nested model was run, and the results are presented in table 4.9.A null model without any predictors was run first. The variance component is relatively small (0.03), with $\chi 2$ (727) = 973.18, p <.001. This translates to an alternative ICC of 0.01, which suggests just 1% of the variation in whether students in the sample chose to major in STEM while in college could be attributed to contextual differences associated with the high schools from which students had graduated. Although high school contexts appear to have little explanatory power in students' choice of a college major, adding school level predictors significantly decreases the variance component by 94.7% percent.

Again, four sets of variables were entered into the model, and the nested-model results are presented in table 4.9. Since research question 2 focuses on identifying the factors, especially whether students undermatched when choosing a college, that relate to their probability of deciding to major in STEM by the beginning of their second year, the first model only contains the variable undermatching status. Note that the matching status used in the descriptive analysis (matched, undermatched, and overmatched) was recoded into undermatching status (undermatched, and not-undermatched) in the HGLM model, the same as the dependent variable in research question 1. The second model adds demographic factors, including gender, racial background, SES status, and language status. The third model adds the school level variables, including school control, urbanicity, and characteristics of high school student body. The final

Table 4.9 Step-wise HGLM Model Predicting Choosing STEM Majors (n = 9050 students, 750 High Schools)

Model	Undermatch				Demographic				High school				Individual background +Zip code			
	Coef.	S.E	Sig.	Δ-p	Coef.	S.E	Sig.	Δ-p	Coef.	S.E	Sig.	Δ-p	Coef.	S.E	Sig.	Δ-p
Matching status																
Academically undermatched	0.396	0.062	***	7.3	0.430	0.065	***	8.0	0.440	0.065	***	8.2	0.091	0.075		
Demographics																
Female					-0.180	0.064	**	-2.8	-0.183	0.064	**	-2.9	-0.198	0.070	**	-3.1
Asian(ref: White)					0.431	0.114	***	8.1	0.373	0.121	**	6.8	0.145	0.128		
Black(ref: White)					0.562	0.104	***	10.8	0.432	0.123	**	8.1	0.599	0.111	***	11.6
Hispanic (ref: White)					-0.042	0.088			-0.112	0.110			-0.031	0.127		
Multiracial (ref: White)					-0.193	0.113			-0.201	0.121			-0.255	0.132		
Lowest quartile SES (ref: Highest quartile)					-0.377	0.156	*	-5.6	-0.390	0.157	*	-5.7	-0.094	0.155		
2nd lowest quartile SES (ref: Highest quartile)					-0.213	0.094	*	-3.3	-0.216	0.110	*	-3.4	-0.159	0.120		
3rd quartile SES (ref: Highest quartile)					-0.234	0.089	**	-3.6	-0.242	0.095	**	-4.3	-0.244	0.102	*	-3.8
First Language English					-0.131	0.081			-0.159	0.084			-0.026	0.087		
High school characteristics																
School control (Ref: public)																
School urbanicity (Ref: urban)									0.120	0.106			0.108	0.124		
% minority									0.012	0.076			0.020	0.080		
Student/teacher ratio									0.005	0.016			0.011	0.023		
Total student enrollment									-0.108	0.096			-0.060	0.110		
% of graduates went to 4-year colleges									0.005	0.006			0.008	0.007		
% of student body is LEP or non-English proficient									0.111	0.039	**	1.9	0.128	0.044	**	2.2
School aggregated SES									0.087	0.045			0.109	0.053	*	1.8
Academic Background																
High school GPA													0.615	0.086	***	12.0
Composite SAT or ACT equivalent score(100)													0.097	0.030	**	1.7
Total AP/IB courses													0.071	0.031	*	1.2
Math self-efficacy													0.194	0.037	***	3.4
Expected level of academic achievement													0.118	0.028	***	2.0
College application																

Number of institutions applied to										-0.046	0.039		
Post-sec school's low expenses important to respondent										0.024	0.056		
Availability of post-sec financial aid important to										-0.038	0.052		
Post-sec school's job placement record important to										0.231	0.055	***	4.1
Whether applied for financial aid										0.323	0.093	***	5.9
College info Personal										0.005	0.031		
College info Formal										-0.026	0.035		
Family influence													
Parents provide advices on academics										0.011	0.036		
Discussion with parents: academics										0.014	0.043		
Friends influence													
# friends plan to attend 4-year college/university										-0.060	0.040		
# friends who consider grades very important										0.009	0.033		
Zip code characteristics													
zip % Bachelor's degree or higher (10%)										0.094	0.047	*	1.6
zip % Speaks English well (10%)										0.024	0.043		
zip % Poverty (10%)										0.071	0.055		
zip % White (10%)										-0.014	0.027		
zip Annual income ($10000)										-0.032	0.022		
Intercept	-1.339	0.031	***	-1.304	0.031	***	-1.289	0.030	***	-1.371	0.034	***	
Variance Component(S.D)	0.019	0.139		0.012	0.109		0.001	0.032		0.020	0.143		
-2 Log Likelihood	27415.7			27358.24			27388.16			27419.86			

model adds other individual level variables, including academic background, college application, family and friends influence, and residential zip code characteristics, depicting the environment of the residential zip code that students come from.

Matching status. As the only predictor in model 1, matching status (whether undermatched or not) does influence students' major choice. Specifically, undermatched students had a higher likelihood of deciding to pursue a STEM major by the spring quarter beginning of their second year in college. This phenomenon is interesting, and further analysis sheds more light on this issue. Undermatched students are more likely to choose STEM, even after demographic and high school variables are controlled. However, when other individual background and zip code variables are added in the final model, undermatched students are not significantly different from their matched or overmatched peers in their probability of majoring in STEM while in college. To further explore which factors are exerting significant influence on this change, significant individual factors are entered one-by-one, on the basis of the third model. Results suggest that the introduction of high school GPA to the model significantly reduces the predictive power of undermatched status. In other words, part of the reason undermatched students selected STEM majors more often than their peers is due to the fact that undermatched students tend to have stronger academic backgrounds; once the model accounted for this fact, differences between undermatched students and their peers were eliminated.

Demographic factors. In the demographic only model, females are 2.8% less likely than males to choose a STEM major, and this difference by gender remains significant after school-level, zip code, and individual background factors are controlled. This shows the similar trend revealed by numerous previous research that women are underrepresented in STEM (Bottia, Stearns, Mickelson, Moller, & Valentino, 2015; Davison, Jew, & Davenport Jr, 2014;

Shapiro & Sax, 2011). Previous research has suggested that pre-college educational settings, secondary school preparation, and interactions with STEM teachers, among other factors, partially account for women's underrepresentation in STEM (Shapiro & Sax, 2011). Given the significant gender effect present even after these other factors are included in the model suggest that women's underrepresentation in STEM persists despite accounting for gender differences in academic achievement, familial backgrounds, and neighborhoods.

In terms of racial background, Asian students tend to choose to major in STEM at significantly higher rates compared to their White peers; however, this difference loses statistical significance in the final model once academic background characteristics and personal background factors. Additional analyses suggest that the additions of GPA, the number of AP/IB courses, and composite SAT scores account for the reduction of the gap between White and Asian students.

Black students are also more likely than White students to choose STEM majors (10.6% higher probability), as shown in the second model (demographics model). However, after controlling for high school variables in the third model, this statistically significant difference was eliminated. Specifically, much of the reason Black students appeared to choose STEM majors in college at higher rates than their White peers was due to the fact that Black students tended to come from more racially diverse high schools, as students more racially diverse high schools tend to major in STEM at higher rates than students from more racially homogenous high schools. The percentage of minority students in high school is the major contributor of this significance reduction in the rate of choosing a STEM major between White and Black students. Interestingly, after individual background and zip code factors entered the model, Black students are again significantly more likely to choose STEM majors. This suppressor effect indicates that

Black students would choose STEM majors at even higher rates than their White Peers if it were not for the fact that Black students in this sample tend to have lower grades and test scores compared to White students.

Still, one might question the result, as prior research (Anderson & Kim, 2006; Musu-Gillette, Robinson, McFarland, et al., 2016) has indicated that minority students are often underrepresented in STEM majors. An important difference between this between this study and previous research pertains to when students' choice of major is being considered. Previous studies have typically considered major choice at the point when students first enter college, likely capturing aspirations for particular academic majors, whereas this study examines choice of major at the beginning of the second year in college. The current study likely also picks up on students' aspirations, as students would need to declare their major by the end of their sophomore year.

With respect to students' SES backgrounds, students from lower SES quartiles have a significantly lower likelihood than those from the highest SES quartile to major in STEM. Specifically, students from the lowest SES and second lowest quartile are 5.6% and 3.3% less likely to decide to major in STEM while in college, but this difference becomes statistically insignificant after individual background and zip code factors enter the model in the final step. Students' standardized test scores largely explain the reduction in predictive power of SES on whether students decide to major in STEM, as students from lower SES backgrounds also tend to score lower on standardized tests relative to their more affluent peers. This result seems to contradict prior research (Niu, 2017; Mullen, 2011; Goyette & Mullen, 2006) that students from lower SES background tend to choose majors that are more directly applicable in the labor market and more lucrative (and STEM majors are often associated with these images). One explanation is that while the aforementioned studies were done on students who have earned a bachelor's degree, the current study sample are students enrolled in all postsecondary institutions, including less selective institutions and community colleges, where the more lucrative job could be things unrelated to STEM at all, such as paralegal. Also, students from the third SES quartile, like their less affluent peers, major in STEM at significantly lower rates than students from the top SES quartile.

Lastly, language status also marginally influences chances of choosing STEM, as students whose first language is English tend to be less likely to major in STEM. Previous research (Chen, 2009) using the 1995-96 Beginning Postsecondary Student Longitudinal data has also identified the same pattern: among students whose first language is not English, 34% entered STEM majors, while among those whose first language is English, only 22% entered STEM majors; however, by the time of graduation, the authors found no differences between these groups in terms of likelihood of completing a STEM degree. In the current study, the effect of language becomes non-significant after high school variables enter the model – specifically once the proportion of LEP (Limited English Proficiency) or non-English proficient student body in high school is controlled for in the model. In essence, non-English speakers tend to come from high schools with higher proportions of non-English speakers, and students of those schools tend to major in STEM more, so this explains why students whose first language is not English are more likely to major in STEM.

High school characteristics. Among all the high school characteristics, only percentage of student body that is LEP and school aggregated SES are significant. Students from schools that have higher percentage of LEP student body tend to have a higher likelihood of majoring in STEM. One possible explanation is that data shows that Asian students tend to attend high schools with higher percentage of LEP (5.3% vs. 11.6% for non-Asian students vs. Asian students respectively), and Asian students are more likely to major in STEM. Additionally, students who graduated from more affluent high schools, which typically offer a broader set of counseling resources and more advanced curricula, also tend to choose STEM disciplines as their major at significantly higher rates than their peers from high schools with lower average SES.

Individual background factors. All five academic background factors are found to be significant, and this is in accordance with previous literature that high school GPA, SAT scores, and math self-efficacy are positively related to the probability of entering STEM fields (Chen, 2009; Moakler & Kim, 2014). Students who completed more rigorous coursework in high school in the form of additional AP/IB classes also are more likely to choose to major in STEM in college (Robinson, 2003; Ehrenberg, 2010). And lastly, students who report aspirations for advanced degrees tend to be more likely to major in a STEM field while in college (Maltese & Tai, 2011; Perez, Cromley, & Kaplan, 2014).

In the college application variables, only having applied for financial aid, and importance placed on postsecondary institutions' job placement record are significant. Students who applied for financial aid are 5.9% more likely than those who did not apply for aid to major in STEM. On one hand, applying for aid may be a proxy for college knowledge/social capital. Those who did not apply for aid may not have known how to navigate that process due to being first-generation lower income, or from less well-resourced schools. On the other hand, students with financial aid have the financial burden, and might be more motivated to major in more directly applicable and lucrative major. The importance placed on career prospective is also found to be positively related to choosing a STEM major, Lastly, one zip code variable, percentage of bachelor's degree or higher in the zip code area is positively related to STEM major choice. Specifically, every 10% increase in the bachelor's degree in the area corresponds with a 1.6% higher probability to choose a STEM major. Another thing to note is that, the percentage of population speaks English well does not influence a student's probability to major in STEM, while as mentioned earlier, the percentage of LEP students in high school does influence the probability to major in STEM.

Influence of undermatching on labor market outcomes

This section presents descriptive and hierarchical linear model analysis on the influence of undermatching on labor market outcomes, and specifically, how does the choice of college major moderate the influence of undermatching on student labor market outcomes. Research question this section aims to answer is:

-*To what extent does undermatching influence students' labor market outcome (i.e. annual earnings from employment)?*

-*To what extent does choice of college major moderate the influence of undermatching on student labor market outcomes (i.e. annual earnings from employment)?*

The dependent variable of research question three is the annual income from employment during 2011(9 years after 10th grade, or 3 years after college graduation). Due to missing value on the variable "number of hours worked per week during 2011" and the variable "employment earnings in 2011", the sample size was reduced to 7220. This study follows previous studies that have examined the relationship between institutional selectivity and earnings (Eide, Hilmer, & Showalter, 2016; Dale & Krueger, 2011; Witteveen & Attewell, 2017), and only selects those working full-time. The sample is thus limited to individuals who 1) have had post-secondary education, regardless of whether they had earned a degree, 2) were working full-time (more than

49 weeks a year, at least 30 hours per week), 3) reported annual income of between $10,000 and $ 200,000 (an arbitrary criteria to eliminate outliers). Therefore, the sample size further reduces to 3860. Consequently, the analytic sample for this research question is different from the previous one (which included all students who were enrolled in postsecondary institutions two years after high school graduation), and caution must be taken while interpreting results, as the analysis only applies to individuals working full-time.

Table 4.10 Proportion of Working status by Alignment between Institutional Selectivity and Academic Eligibility, Weighted (n=7220)

	Not undermatched (n=4440)	Undermatched (n=2780)	Total
	%	%	%
Not working full-time (n=3230)	44.9	42.6	44.0
Working full-time (n=4000)	55.1	57.4	56.0

* Pearson chi2(1) = 3.6881 Pr = 0.128

Table 4.10 presents the proportion of individual working full-time across undermatched and not-undermatched individuals. Working full-time was defined as working 30 hours a week or more, and worked more than 49 weeks in 2011. The table shows that in the sample, the majority was working full-time (56.0%), and around two out of five (44.0%) were either working part-time, or not working. It looks like among undermatched and not-undermatched population, there is almost no difference in the proportion of people who worked full-time, and the chi-square test confirms the result (p>05).

Table 4.11 listed the descriptive statistics of the new sample, including existing variables used in previous research questions, and new variables added. Compared to the full sample that included individuals who were enrolled in a postsecondary institution two years after high school graduation, this analytic sample is slightly different, but most of the changes are minimal (less than difference of .03 in mean). However, it is worth noting that in this analytic sample, the percentage of undermatched students is 0.40, while the original sample has an undermatching rate of 0.36, showing that the new sample has more undermatched students. Thus, it appears that students who undermatched were more likely to be employed full-time in 2011 than those who did not undermatch when selecting their college. Meanwhile, the original sample has 21.1% choosing a STEM major two years after high school graduation, and among the new sample the percentage is 23.9% (not listed in table 4.11), suggesting that STEM aspirants are slightly more likely to be employed full-time in 2011. Moreover, in the new sample, 15.0% ultimately obtained a credential in STEM.

Regarding the new variables involved in the new sample, about 30% are married. Roughly one in three (33%) has a professional certification or license, and a majority (52%) have ever received formal employer-provided training. As to the credential level, slightly over half (52%) had enrolled in college since graduating from high school but had not earned a credential, 11% had earned an associate's degree, an undergraduate certificate, or diploma, and approximately two out of five (39%) had obtained a bachelor's degree or higher. In this analytic sample, 19% were academically eligible for *very selective* institutions, yet just 13% attended one; 28% were eligible for enrollment at *selective* institutions, but only 19% attended an institution of that selectivity level. Additionally, 27% attended a *somewhat selective* institution, and almost two out of five (41%) attended a *nonselective or two-year* institution. Lastly, about one out of four (25%) attended a public institution, 2% attended for-profit institution and HBCU institutions.

Table 4.11 Description of Student-level and High School Level Variables in HLM models (weighted n= 3860 students, 710 high schools)

	Min	Max	Mean	Std. Dev.
Existing Variables				
Demographics				
Female	0.00	1.00	0.54	0.50
Asian(ref: White)	0.00	1.00	0.08	0.27
Black(ref: White)	0.00	1.00	0.09	0.29
Hispanic (ref: White)	0.00	1.00	0.09	0.29
Multiracial (ref: White)	0.00	1.00	0.04	0.19

Lowest quartile SES (ref: Highest quartile)	0.00	1.00	0.13	0.34
2nd lowest quartile SES (ref: Highest quartile)	0.00	1.00	0.20	0.40
3rd quartile SES (ref: Highest quartile)	0.00	1.00	0.28	0.45
First Language English	0.00	1.00	0.88	0.32
Academic background				
High school GPA	0.71	4.20	3.09	0.60
Composite SAT or ACT equivalent score(100)	4.50	16.00	10.27	1.92
Total AP/IB courses	0.00	18.00	1.16	1.93
Math self-efficacy	-2.04	1.85	0.12	0.91
Expected level of academic achievement	2.00	8.00	6.75	1.15
College application				
Number of institutions applied to	0.00	4.13	2.70	0.90
Post-sec school’s low expenses important to respondent	1.00	3.00	2.11	0.68
Availability of post-sec financial aid important to respondent	1.00	3.00	2.39	0.71
Post-sec school’s job placement record important to respondent	1.00	3.00	2.51	0.63
Whether applied for financial aid	0.00	1.00	0.77	0.42
College info Personal	0.00	4.00	1.80	1.09
College info Formal	0.00	7.00	2.24	1.10
Family influence				
Parents provide advice on academics	-1.81	1.29	0.10	0.88
Discussion with parents: academics	-2.70	1.52	0.19	0.83
Friends' influence				
# friends plan to attend 4-year college/university	1.00	5.00	3.71	0.90
# friends who consider grades very important	0.00	3.00	1.44	0.99
Residential zip code influence				
Zip % Bachelor's degree or higher (10%)	0.00	10.00	2.93	0.84
Zip % Speaks English well (10%)	3.66	10.00	9.31	0.91
Zip % Poverty (10%)	0.00	6.69	2.72	0.74
Zip % White (10%)	0.14	10.00	7.91	2.16
Zip Annual household income ($10000)	2.48	29.05	7.50	3.00
High School level predictors				
High School characteristics				
School control (Ref: public)	0.00	1.00	0.23	0.42
School urbanicity (Ref: urban)	0.00	1.00	0.68	0.47
% minority	0.00	10.00	3.02	3.08
Student/teacher ratio	0.15	4.48	1.67	0.48
Total student enrollment	0.00	44.00	12.30	7.95
% of graduates went to 4-year colleges	1.00	6.00	4.51	1.14
% of student body is LEP or non-English proficient	0.00	5.00	0.54	0.91
School aggregated SES	-0.81	1.40	0.14	0.38
New variables				
Credential in STEM	0.00	1.00	0.15	0.36
Undermatched	0.00	1.00	0.40	0.49
Married	0.00	1.00	0.30	0.46
Has a professional certification or license	0.00	1.00	0.33	0.47
Received formal employer-provided training	0.00	1.00	0.52	0.50
Credential level: No credential	0.00	1.00	0.52	0.50
Credential level: Associate's degree or UG certificate/diploma	0.00	1.00	0.09	0.29

Credential level: Bachelor's degree or higher	0.00	1.00	0.39	0.49
Academically eligible for: very selective	0.00	1.00	0.19	0.40
Academically eligible for: selective	0.00	1.00	0.28	0.45
Academically eligible for: somewhat selective	0.00	1.00	0.25	0.43
Academically eligible for: nonselective/2-year institutions	0.00	1.00	0.28	0.45
First attended: very selective	0.00	1.00	0.13	0.33
First attended: selective	0.00	1.00	0.19	0.39
First attended: somewhat selective	0.00	1.00	0.27	0.44
First attended: nonselective/2-year institutions	0.00	1.00	0.41	0.49
First PSI control (Ref: public)	0.00	1.00	0.25	0.43
First PSI sector (Ref: not-for profit)	0.00	1.00	0.02	0.13
First PSI HBCU	0.00	1.00	0.02	0.15
First PSI Total enrollment (100)	0.32	922.81	177.30	146.44
First PSI % White (10%)	0.00	9.82	6.69	2.08
First PSI % Female (10%)	0.00	10.00	5.64	0.96
First PSI Log tuition and fees	6.25	11.04	8.66	0.94

Table 4.12 presents the average annual income of undermatched and not-undermatched individuals that were working full time across different categories, including gender, racial background, language status, socioeconomic status, and academic background, and two school level categories. Statistics are obtained after Taylor Series Linearization, in order to account for the non-Simple Random Sample design in the ELS 2002 questionnaires.

Table 4.12 Summary of Annual income by Subgroups for weighted national sample (n = 3860)

	Not undermatched (n=2300)		Undermatched (n=1560)		p-value	Sig.
	Mean($)	SD($)	Mean($)	SD($)		
Overall (n=3860)	39315	888	38030	595	0.237	
Individual						
Degree status						
Didn't obtain any certificate (n=1990)	34710	671	35920	884	0.2882	
UG certificate or diploma, or associate's (n=360)	35175	1659	34325	1226	0.6785	
Bachelor's degree or higher (n=1510)	47107	2129	42470	843	0.0474	*
Credential field						
Obtained STEM credential (n=570)	45640	1805	47966	1374	0.3145	
Did not obtain a STEM credential (n=3290)	38432	955	36138	627	0.0474	*
Gender						
Male (n=1790)	43671	1699	42045	1068	0.418	
Female (n=2070)	35119	696	34975	628	0.8783	
Racial background						
Asian (n=300)	46298	2616	51421	3971	0.2808	
Black (n=360)	31977	1242	37679	3529	0.1287	
Hispanic (n=360)	34980	1160	34050	1946	0.6946	
Multiracial (n=150)	39648	4529	37514	4481	0.7401	
White (n=2690)	41220	1259	37990	614	0.0237	*

SES						
SES lowest quartile (n=510)	33285	1085	33821	1443	0.7759	
SES second lowest quartile (n=770)	38674	3462	34743	974	0.2765	
SES third quartile (n=1070)	37026	945	38798	965	0.2076	
SES highest quartile (n=1510)	44045	1134	41169	1149	0.0812	
Language status						
English not as the first language (n=450)	46666	6006	39703	2412	0.2889	
English as the first language (n=3410)	38377	676	37931	599	0.6251	
Academic eligibility						
Very selective (n= 750)	53930	2503	42925	1167	0.0001	***
Selective (n=1060)	44106	1153	38406	901	0.0002	***
Somewhat selective (n= 950)	38093	1042	34330	1148	0.0106	*
Nonselective or 2-year (n= 1100)	34566	1582	32558	2278	0.4685	
First institution selectivity						
Very selective (n=480)	50516	1716	~	~		
Selective (n=740)	42543	1285	43412	1906	0.7117	
Somewhat selective (n=1040)	37840	2692	39536	895	0.5492	
Nonselective or 2-year (n=1600)	33242	913	36102	809	0.025	*
High school						
High school control						
High school public (n=2700)	38647	998	37631	635	0.3986	
High school private (n=1160)	44256	1087	41861	1592	0.2104	
High school urbanicity						
High school urban (n=1260)	38405	1038	38623	1365	0.9081	
High school non-urban (n=2600)	39702	1184	37867	658	0.1755	

*$p<.05$, **$p<.01$, ***$p<.001$

Firstly, Table 4.12 tells the general earning situation across different groups. Judged from the overall undermatched and not-undermatched annual income comparison, it seems that there is not much difference. The average for those undermatched is $38,030, and those not undermatched is $39,315, and the F-test shows that the difference is not significant. Moreover, among all the variables listed, degree status, credential field, racial background, academic eligibility, and first attended institution selectivity categories see significant differences among students who undermatched, and students who matched and over-matched.

Firstly among bachelor's degree earners, on average undermatched individuals earn $42,470, while those not undermatched earn $47,107 (gap of $4,623). Moreover, the difference between undermatched ($36,138) and not-undermatched ($38,432) individuals who had not obtained a STEM credential is also significant (gap of $ 2,294).

Among measures related to students' academic strengths, those who were academically eligible for *very selective* institutions have the highest annual earnings among the four groups. The difference between undermatched ($42,925) and not-undermatched ($53,930) is also the largest among the four comparison groups (gap of $11,005), and this difference is significant ($p<.001$). Among students who were academically eligible for *selective* institutions, students that were not-undermatched have an average annual earnings of $44,106 but for students that were undermatched the number is $38,406 ($5,700 lower), and this difference is also significant at $p<.001$ level. The same pattern could be observed among students who were academically eligible for *somewhat selective* institutions ($p<.05$). Therefore, it looks like the negative effect of undermatching is obvious, but this descriptive finding needs to be further explored by also controlling for other covariates.

Considering earnings differences by students' undermatched status broken out by levels of institutional selectivity tells another story. The general pattern is that for students who attended similarly selective institutions, undermatched students earn significantly more than those not-undermatched. By definition, students who undermatched had stronger academic backgrounds than their peers who had enrolled in the same institution, and, as a result, they earned slightly more than their peers in 2011; however, most of the differences are not significant ($p>.05$). The only significant difference ($p<.05$) lies in the *first attended nonselective or 2-year institution* category, with those not undermatched having earned $33,232 and those undermatched having earned $36,102 ($2,870 higher). This result suggested that undermatched students become "big frog in a small pond" (Davis, 1966, p.31).

HLM models: Interaction between undermatching status and credential field on labor market outcomes

Table 4.13 presents the HLM models results predicting log employment earnings nine years after 10th grade. Model 1 only includes the core terms- undermatching status, credential field, and the interaction term. Model 2 includes background factors, including demographics, high school characteristics, zip code area characteristics, and training. Model 3 further includes other key formal education factors, and Model 4 adds post-secondary institution characteristics. The final multilevel model explains 20.7% variance in earnings (a noticeable increase from 1.3% in the initial model) at student level, and this final model explains 27.7% variance among high schools.

Model 1 results show that when only considering undermatching status and credential field, no significant difference in earnings exists between students who undermatched in their college choice and those whose choice suggested a good match or overmatch. Credential field, however, does impact employment earnings. Specifically, among students who are not undermatched (the reference group) , those getting credentials in STEM field on average would earn 21.7% more than those not. Previous research also found the influence of major on earnings, and that STEM majors earn more than non-STEM majors (Eide, Hilmer, & Showalter, 2016). The interaction term between credential in STEM and undermatching status is also significant, and by combining the coefficient of undermatching status and the interaction term, it could be seen that among STEM credential holders, those undermatched earn 8.4% (calculated from exp (0.109-0.028)) more than those not-undermatched; among those without a STEM credential however, those undermatched earn slightly less but almost the same as those not-undermatched.

The second model adds background information, including demographics, high school characteristics, zip code area characteristics, and training. The coefficients for the three initial variables remain stable and significant after these additional covariates are included, and several other variables emerged as significant. Firstly, keeping everything else constant, females earn significantly less (15.0%) than males. The influence remains strong even after formal education and postsecondary institution variables enter the model. Referring back to the descriptive table 4.12 at the beginning of this section (the differences between male and female in earnings, for both undermatched and not-undermatch, are quite large), it looks like after controlling for other variables, the gender gap remains large (15.1%).

Among the racial group variables, Black and Hispanic individuals earn significantly lower annual incomes compared to their White peers. Specifically, Black students would earn 8.7% less than White students, and Hispanic students would earn 10.8% less than White students. All the SES group variables are significant ($p<.001$), and keeping everything else constant, students from the lowest quartile SES earned 16.8% less than the highest quartile SES students; students from the second lowest quartile earned 12.2% less than the students from the most affluent families; and students from the third quartile earned 7.9% less than their peers in the top quartile. However, adding more variables does change the significance level of these variables, and further explanation is provided in later paragraphs.

The influence of language status is similar to the trend depicted in the descriptive results in table 4.12. Specifically, keeping everything else constant, students whose first language is English earned significantly less than those English-not-as-first-language students, and the difference is 15.0%. Lastly, being married also significantly influences annual income: those married on average earned 15.9% more than those not married (including never married,

Table 4.13 HLM Models Predicting Log Employment Earnings 9 Years After 10th Grade (N=3860 students, 710 high schools)

	Basic			Individual background			Formal education			PSI		
	Coef	exp(b)	Sig.	Coef	exp(b)	Sig.	Coef	exp(b)	Sig.	Coef	exp(b)	Sig.
Academically undermatched	-0.028	0.972		-0.019	0.982		-0.102	0.903	***	-0.080	0.923	**
Credential in STEM	0.197	1.217	***	0.130	1.138	***	0.056	1.058		0.055	1.057	
Credential in STEM*undermatched	0.109	1.115	*	0.119	1.127	*	0.135	1.145	**	0.137	1.147	**
Demographics												
Female				-0.146	0.864	***	-0.166	0.847	***	-0.164	0.849	***
Asian(ref: White)				0.042	1.043		0.007	1.007		-0.006	0.994	
Black(ref: White)				-0.091	0.913	*	-0.044	0.957		-0.026	0.974	
Hispanic (ref: White)				-0.115	0.892	*	-0.095	0.910	*	-0.108	0.897	*
Multiracial (ref: White)				-0.094	0.910		-0.096	0.909		-0.099	0.906	
Lowest quartile SES (ref: Highest quartile)				-0.183	0.832	***	-0.116	0.890	***	-0.113	0.894	***
2nd lowest quartile SES (ref: Highest quartile)				-0.130	0.878	***	-0.068	0.934	*	-0.069	0.933	*
3rd quartile SES (ref: Highest quartile)				-0.082	0.921	***	-0.047	0.954	*	-0.048	0.953	*
First Language English				-0.163	0.850	**	-0.174	0.841	**	-0.168	0.846	**
Married				0.148	1.159	***	0.154	1.166	***	0.156	1.169	***
Pre-labor market influences: high school characteristics												
% of graduates went to 4-year colleges				0.016	1.016		0.005	1.005		0.003	1.003	
School control (Ref: public)				0.067	1.069	*	0.054	1.056		0.038	1.039	
School urbanicity (Ref: urban)				-0.005	0.995		-0.001	0.999		-0.001	0.999	
School aggregated SES				-0.063	0.939		-0.078	0.925		-0.082	0.922	*
% minority				-0.005	0.995		-0.007	0.993		-0.009	0.991	
Student/teacher ratio				-0.010	0.990		-0.018	0.982		-0.022	0.978	
Total student enrollment				0.001	1.001		0.002	1.002		0.002	1.002	
% of student body is LEP or non-English proficient				0.005	1.005		0.004	1.004		0.006	1.006	
Pre-labor market influences: zip code area												
zip % Bachelor's degree or higher (10%)				0.007	1.007		0.009	1.009		0.005	1.005	
zip % Speaks English well (10%)				-0.026	0.975		-0.028	0.972		-0.020	0.981	
zip % Poverty (10%)				0.012	1.012		0.012	1.012		0.009	1.009	
zip % White (10%)				-0.001	0.999		-0.005	0.995		-0.007	0.993	
zip Log annual income				0.174	1.190	***	0.131	1.140	***	0.124	1.132	**

Training												
Has a professional certification				0.099	1.104	***	0.100	1.106	***	0.099	1.104	***
Has received formal employer-provided training				0.088	1.092	***	0.084	1.088	***	0.084	1.088	***
Formal education: other factors												
UG certificate or diploma, or associate's (ref: no certificate)							-0.012	0.988		-0.011	0.989	
Bachelor's degree or higher(ref: no certificate)							0.095	1.100	***	0.084	1.088	***
AE very selective (Ref: nonselective and 2-year)							0.280	1.323	***	0.231	1.260	***
AE selective (Ref: nonselective and 2-year)							0.204	1.226	***	0.177	1.194	***
AE somewhat selective (Ref: nonselective and 2-year)							0.110	1.116	***	0.097	1.102	***
Formal education: First PSI characteristics												
First PSI Total enrollment (100)										0.000	1.000	
First PSI control (Ref: public)										0.019	1.020	
First PSI sector (Ref: not-for profit)										-0.095	0.909	
First PSI HBCU										-0.205	0.815	**
First PSI % White (10%)										-0.015	0.985	*
First PSI % Female (10%)										-0.010	0.990	
First PSI Log tuition and fees										0.025	1.026	
Intercept	10.436	34068	***	10.456	34760	***	10.461	34910	***	10.460	34889	***
High school variance components (S.D)	0.00033	0.01813		0.00188	0.04339		0.00002	0.00408		0.00002	0.00424	
Level-1 Variance components (S.D)	0.23766	0.4875		0.2023	0.44978		0.19385	0.44028		0.19099	0.43702	
Deviance	5416.3			4833.7			4621.6			4564.3		
Variance accounted for by L1	0.013			0.153			0.196			0.207		
Variance accounted for by L2	0.096			0.209			0.266			0.277		

divorced, separated, and widowed).

Among high school characteristics factors, only one variable was significant at the $p<.01$ level: students from private high schools earned 6.9% more than students from public high schools, keeping everything else constant. Among all the zip code level variables, a strong effect on earnings came from the average income of the communities where students lived prior to college. As both the dependent variable (annual earnings) and the variable zip code annual income are log-transformed, the interpretation is different from above, and the coefficient suggests that a 1% increase in the average annual income of a zip code is associated with a 0.124% increase in estimated annual earnings.

In terms of training, professional certification significantly influences employment income, and obtaining a professional certification increased annual earnings by 10.4%. This result further confirms positive effect of professional certification on earnings, as revealed by previous literature (Wiley, 1995; Weeden, 2002). Meanwhile, receiving formal employer-provided training also significantly increases earnings, and those receiving trainings would earn 9.2% more than those who did not. This also confirms previous findings that employer provided training is associated with higher earnings (Barron, Black, & Loewenstein, 1987; Lynch, 1992).

Model 3 adds other formal education factors, and it could be seen that both academic eligibility and credential level significantly correlate with annual income. On average, keeping everything else constant, respondents who were academically eligible for admission to and enrollment at *very selective* institutions earned 32.3% more than their counterparts eligible for *nonselective and 2-year* institutions; students eligible for *selective* institutions earned 22.6% more; and students eligible for *somewhat selective* institutions earned 11.6% more than those eligible to enroll at *nonselective and 2-year* institutions. Additionally, students obtaining a bachelor's degree reported earnings 10.0% more than those not getting a certificate, while getting undergraduate certificate or diploma, or getting an associate's degree does not seem to give an edge relative to individuals who did not earn any type of postsecondary credential.

Adding academic eligibility and credential status also changes previous coefficients to a great extent. Specifically, being undermatched jumped from being non-significant to significant at $p<.001$ level. Furthermore, controlling for everything else, undermatched students earned 9.7% less than their peers who went to an academically matched or more selective institution. The suppressor effect suggests undermatched students would have earned even less if it were not for the fact that they tend to be those more academically suited for higher selectivity universities. In essence, once the model controlled for the strength of students' academic eligibility and the earnings advantage that eligibility for enrollment at more selective institution entails, undermatched students were at a disadvantage with respect to annual incomes when compared to their matched or overmatched peers with the same academic eligibility. The interaction term influence becomes more obvious, but getting a credential in STEM fell out of significance. One way to interpret such finding is that, students who earned a credential in STEM are also those who are more likely to obtain a bachelor's degree, and those that are academically eligible for more selective institutions.

Another finding to note is that adding other formal education variables to the model eliminates some of the significant differences associated with demographic characteristics in the first three models. For instance, the inclusion of other educational measures eliminated the significant earnings gap between Black individuals and their White peers, which occurs once the model controls for the type of credential earned by respondents, suggesting that the racial earnings gap was largely attributed to the advantage White students have over their Black peers with respect to their likelihood of completing a bachelor's degree. By contrast, the earnings gap between White non-Hispanic and Hispanic respondents persists even after controlling for these other educational factors. The addition of these academic-related characteristics partly explains away the previously identified gap between respondents in the highest SES quartile and their peers in lowest and second lowest SES quartiles, but the gap between the third quartile and the top quartile persists in the final model. This finding reveals that students from the third SES quartile face other obstacles that prevent them from acquiring higher earnings, aside from the obstacles in academic eligibility, credential status, and contextual influences of their high school and college.

Lastly, the only college-level characteristics significantly associated with annual earnings were an indicator related to whether students' first institution they attended was designated as an HBCU and the percentage of White student on campus. Noticeably, students in HBCU institutions earn 18.5% less than those attending other institutions. Moreover, it looks like a more diverse campus also benefit students in terms of earnings, and specifically, every 10 percent increase in the White population on campus translates to 1.5% decrease in annual income.

More importantly, adding first attended institution characteristics also decreases the strength of the undermatching variable. And in the final model, it could be observed that among students who obtained a STEM credential (regardless of whether it is a bachelor's degree, associate's degree, diploma or certificate) undermatched students earn 5.9% (calculated from exp (0.137-0.08)-1) more than those not undermatched; however, among those getting a non-STEM credential or not getting any certificate, undermatched students earn 7.7% less than their not-undermatched peers. Therefore, the final conclusion is that majoring in STEM significantly moderates the negative impact of undermatching on earnings.

Subgroup Earning Analysis

One limitation of table 4.13 is that it could not control for academic eligibility, first attended institution, and undermatching status all together, due to multicolinearity issue (undermatching status was operationalized by comparing academic eligibility level and first attended institution selectivity level). As a result, models in table 4.13 only controlled for undermatching status and academic eligibility level. Therefore, the influences of undermatching and first attended institution selectivity level are still somewhat entangled. To further breakdown the influence of undermatching, the undermatching status variable was removed; instead, separate nested models were created for students of different academic eligibility level, and the reference groups of *first attended institution selectivity* were changed according to academic eligibility level. In this way, the significance levels of core term coefficients signal the effect of attending a not-matched institution on earnings.

Table 4.14 presents the results of HLM models that predict log employment of students, separated by their academic eligibility level. For students of each academic eligibility level, two models were created –the core model only includes *getting a certificate in STEM*, *first attended institution selectivity level*, and the interaction terms, and the full model controls for all background variables identified in table 4.13, including demographics, high school characteristics, zip code area characteristics, training, credential level, an first postsecondary institution characteristics. As the focus of this section is to compare the coefficient of core terms across academic eligibility levels, and how they changed once controlling for all background covariates, Table 4.14 only presents the core terms in first and full model, and full nested models could be seen in appendix D.

Table 4.14. HLM Models Predicting Log Employment Earnings 9 Years After 10th Grade, by Academic Eligibility Level, comparison between core term model and full model (N=3860 students, 710 high schools)

	Academically eligible for											
	Very selective (n= 750)						Selective (n=1060)					
	Core model			Full model			Core model			Full model		
	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.
Credential: STEM field	0.089	1.093		-0.142	0.868		0.219	1.245	**	0.189	1.208	*
FI very selective	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	0.098	1.103		0.03	1.03	
FI selective	-0.254	0.776	***	-0.121	0.886		(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
FI somewhat selective	-0.308	0.735	***	-0.096	0.909		-0.104	0.901	*	-0.05	0.948	
FI non-selective or 2 year	-0.15	0.861	*	0.146	1.158		-0.125	0.882	*	0.02	1.02	
INT: FI very selective*STEM	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	0.063	1.065		0.068	1.07	
INT: FI selective*STEM	0.325	1.385	**	0.419	1.521	**	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
INT: FI somewhat selective*STEM	0.261	1.298	*	0.387	1.472	**	0.057	1.059		-0.01	0.992	
INT: FI non-selective or 2 year *STEM	0.122	1.13		0.228	1.256		0.001	1.001		0.02	1.02	
Intercept	10.667	42917	***	10.665	42833	***	10.53	37407	***	10.53	37508	***

	Academically eligible for											
	Somewhat selective (n=950)						Nonselective and 2-years (n=1100)					
	Core model			Full model			Core model			Full model		
	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.
Credential: STEM field	0.152	1.164		0.012	1.012		0.1	1.106		-0	0.996	
FI very selective	-0.012	0.988		-0.061	0.94		0.141	1.151		0.034	1.035	
FI selective	0.161	1.175	*	0.079	1.083		0.002	1.002		-0.03	0.967	
FI somewhat selective	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	0.053	1.054		0.033	1.034	
FI non-selective or 2 year	-0.111	0.895	*	-0.066	0.936		(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
INT: FI very selective*STEM	0.056	1.057		0.182	1.199		-0.316	0.729		-0.27	0.76	
INT: FI selective*STEM	-0.07	0.933		-0.013	0.987		0.073	1.075		0.05	1.051	
INT: FI somewhat selective*STEM	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	0.256	1.292	*	0.313	1.368	**
INT: FI non-selective or 2 year *STEM	0.114	1.121		0.261	1.298		(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
Intercept	10.393	32639	***	10.404	32990	***	10.293	29528	***	10.29	29327	***

*FI= First attended Institution

Firstly, among students whose academic qualifications are best suited for *very selective* institutions, in the core model, students attending less selective institutions did earn significantly less. Specifically, compared to their peers who first attended *very selective* institution, students first attending a *selective* institution earn 22.4% less, and the number is even larger for those first attending a *somewhat selective* institution (26.5%). Interestingly, those first attended *non-selective or 2-year* institutions also earned less, but the difference is smaller (13.9%). However, after controlling for other background variables, all *first attended institution selectivity* terms fell out of significance. Additional analysis shows that it is the entrance of postsecondary institution characteristic variables into the model that prompted this change, and two salient variables are postsecondary institution sectors (for-profit or not), and percentage of White students on campus. Specifically, students in for-profit institutions earn 44% less, and every 10 percent increase in White population means 4.6% less in annual income.

While getting a credential in STEM did not increase earnings for students who are academically eligible for *very selective* institutions, two of the interaction terms are significant, even after other variables are controlled for. Therefore, students who first attended a *selective* institution but got a credential in STEM earn 31.9% (calculated from exp(-0.142+0.419)-1) more than their peers who also attended a *selective* institution but did not got a STEM credential. So importantly, earning a STEM credential reverses the disadvantage of being undermatched for those eligible for enrollment at very selective institutions. In other words, perhaps it can benefit the highest achieving students to attend an institution of a lower selectivity level if they intend to major in STEM. This could potentially allow them to be a bigger frog in a smaller pond and thus find easier access to research opportunities, or be leaders in the classroom whereas those opportunities may have been harder to come by if they had been in more competitive environments. Furthermore, among students who attended a *somewhat selective* institution, those with a STEM credential earn 27.8% more, and among students who attended a *nonselective* or 2-year institutions, the earnings of those with a STEM credential does not differ from those without one.

Now switch to students who are academically eligible for *selective* institutions. The positive effect of getting a credential in STEM still persists in the full model, and in general, students who got a credential in STEM earn 20.8% more than those with a credential in non-STEM or not have a credential. In the initial model, students who attended *somewhat selective* or *nonselective/2-year* institutions earned significantly less (9.9% and 11.8% respectively) than their peers who attended an institution that matched with their academic credential, but these effects disappear after controlling for other variables. Additional analysis shows that adding the community economic status variable is the reason of change. Specifically, every one percent change in zip code annual average household income translates to 0.15% increase in individual earnings. This also shows that students attending less selective institutions than their academic qualifications are more likely to come from zip codes with lower household income. None of the interaction terms are significant, suggesting that across students who first attended different selectivity institution, getting STEM credential always increases income.

Among students who are academically eligible for *somewhat selective* institutions, almost no pattern could be observed. The initial model shows that students attending a *selective* institution earn 17.5% more than their peers that attended a *somewhat selective* institution, and respectively students attending a *non-selective or 2-year* institution earn 10.5% less. However, this effect shifts to being non-significant in the full model. Similarly, for students who are academically eligible for *nonselective/2-year* institutions only, none of the coefficients are significant, suggesting that factors other than credential field or higher education institution selectivity are influencing their earnings.

Summary. This section presents results firstly related to predictors of undermatching status. Specifically, in addition to racial background, parental SES factors, student college choice priorities on postsecondary institution's low expense, availability of financial aid, and job placement are found to influence undermatching. College information sources however, are not found to be significant. Students' community socioeconomic environment, represented by zip code annual household income also influences undermatching status. As to the predictors of choosing a STEM major, it was found that among undermatched population a higher percentage of students chose STEM, compared to those attending a matched or even more selective institution. However, this difference could be explained by their academic qualifications.

Lastly, the analysis on earnings shows that majoring in STEM does significantly moderate the negative impact of undermatching on earnings but only for students eligible to enroll in very selective institutions. Interestingly, among students who earned a non-STEM degree or did not earn a credential, being undermatched significantly decrease their earnings; however, among students who obtained a credential in STEM, undermatched student actually earn more. This effect is also the most salient among students who have the best academic qualifications. Chapter 5 will continue to discuss the implication of such findings.

CHAPTER 5: DISCUSSION AND CONCLUSION

The final chapter presents a brief overview of this study's objectives and the three research questions, reviews the conceptual framework that guided the design of study, as well as the methodology and analytic techniques used. Discussion of findings, implications for K-12 and higher education are also provided, along with limitation and direction for future research.

Overview of the Study

Traditionally, colleges and universities have been expected to promote social mobility (Haveman & Smeeding, 2006). It has been widely recognized that higher education is one of the best investments an individual can make. With the current situation that 66% of high school graduates enroll in postsecondary institutions (Bureau of Labor Statistics, 2014), the hierarchical structure within higher education system increasingly exerts a crucial influence on the distribution of life chances. Greater focus now has been placed upon where individuals actually went to college, instead of simply whether one went to college or not. The relationship between college selectivity and earnings has been demonstrated by the fact that higher selectivity is generally associated with higher earnings (Hoekstra, 2009; Beyond, Brewer, Eide , & Ehrenberg,1999). Hence, it can be argued that when highly qualified students from low-income backgrounds undermatch, they ultimately may limit the potential for higher education to be a force in support of social mobility.

In addition to the fact that earnings are associated with college selectivity, the major field of study students choose is also influential. As a factor that has long been recognized, college major exerts great influence on college graduates' labor market outcomes (Rumberger &Thomas, 1993; Thomas 2003). However, there lacks empirical studies that explores the influence of postsecondary undermatching on students' labor market outcomes, and especially the different influence of undermatching in different academic field (STEM and non-STEM). Therefore it is essential to understand the role college major plays when studying the effect of undermatching on students' labor market outcomes.

Therefore, this study examines *who, how* and *what* of the relationship between undermatching and choosing a STEM major. The design of this study was guided by two sets of conceptual framework, including the college decision framework adapted from Perna (2006) and Iloh (2018), and human capital theory (Becker, 1975; Mincer, 1957). The college decision model considers the influence of individual, community and high school characteristics on college decisions, including where to go, and what majors to choose. The framework based on human capital theory (Becker, 1975; Mincer, 1957) focuses on how schooling, training, and pre-labor market environment influences individual's income.

The study utilized data from three sources, including the Educational Longitudinal Study of 2002 (ELS 2002), American Community Survey (ACS 2005), and Integrated Postsecondary Data System (IPEDS). The ELS 2002 is a nationally representative, longitudinal study of 10th graders in 2002, and followed through their secondary and postsecondary years, and includes variables such as students' demographic backgrounds, family backgrounds, high school experiences, college application and choices, college experience, and labor market outcomes. ACS 2005 provides zip code level data regarding information about ancestry, educational attainment, income, language proficiency and employment. IPEDS provides postsecondary institution characteristics information such as study body composition, and tuition and fees.

Two sets of quantitative analysis were conducted for each research question. For the first research question, cross-tabulations with chi-squared test were first used to compare the academic eligibility level across racial background and SES quartiles. Additionally, proportion tables were created to reflect the proportion of matching status (undermatched, matched, overmatched) across specific combinations of academic eligibility level and racial background/SES quartiles. HGLM was then used to understand what factors influence students' chances of undermatching in the context of high school and residential area. For the second research question, processes were almost the same. Proportion tables were created to reflect the proportion of choosing STEM across specific combinations of undermatching status and racial background/SES quartiles, then HGLM was used to understand how do undermatching status and other factors influence students' choice of STEM major in the context of high school and residential area. The sample for research question 1 and 2 were limited to those who were still enrolled in postsecondary two years after high school graduation, and the sample size is 9050.

For the third research question, descriptive analysis first examined how individual's employment earnings vary by undermatching status and background factors (eg. race, gender, SES, academic background, credential status), along with F-test to test significant differences in the average earnings across undermatching status in the population. Then HLM was then used to explore the extent to which undermatching status and earning a STEM postsecondary credential influence earnings. The sample is limited to those who had enrolled in a college or university by 2006 and who reported working full-time (more than 30 hours a week, more than 49 weeks a year in 2011), which resulted in an analytic sample of 3860.

Discussion of Findings

Research Question One: Undermatching status

I hypothesized that on individual level, racial background, parental SES, students' college choice priorities, and the number of college information sources would significantly influence students' chances of undermatching. I also hypothesized that zip code area level socioeconomic indicators would influence individual's undermatching status.

The descriptive and multilevel analysis results indicate that some of the hypothesis above holds true, while also revealing more complicated facts about undermatching. In terms of racial background, one of the most interesting finding is the undermatching reality of Black students. It was found that in the weighted national sample, White students undermatched at the highest rate (44.4%), while Black students undermatched at the lowest rate (12.4%). The undermatching rate for Asian, Hispanic, and Multiracial students are 30.1%, 25.1%, and 30.7% respectively. This contradicts previous findings that there were higher undermatch rate among minority students, especially Black and Hispanic students (Bastedo & Jaquette, 2011; Roderick, Nagaoka, Coca & Moeller, 2008).

However, the above-mentioned low undermatch rate among Black students could be quite misleading, if not considering the rate of undermatch at each selectivity level. Take Black students as example, even though their overall undermatch rate is the lowest (12.4%), and far below the national average (36.9%), the undermatch rate is the highest (74.8%) among students who are academically eligible for very selective institutions (compared to the national average of 61.6%), and the second highest (61.5%) among students who are academically eligible for selective institutions (compared to national average of 60.2%). In contrast, Black students undermatch rates are at the lower end among students who are academically eligible for somewhat selective and nonselective institutions. Essentially, among the most academically talented students, Black students are the least likely to enroll at a college or university with a selectivity level that matches their eligibility.

Additionally, the overmatch rate of Black students follows the same trend, with Black students eligible for enrollment at *selective* institutions among the least likely to enroll at a *very selective* college or university. Importantly, overmatch does not mean that students are under-qualified; institutions exercise holistic review to best meet institutional goals, and they consider factors such as athletic ability, artistic talent, race, gender, legacy status, economically disadvantaged background, and personal experience (Espenshade, Chung, & Walling, 2004; Fetter, 1997). Additionally, higher education institutions admit particular students, believing that they would thrive here. The overall overmatch rate of Black students is 34.3%, far beyond the national average of 17.6%. While this looks promising, further selectivity breakdown tells another story. Among Black students who are academic eligible for selective institutions, only 12.4% overmatch, the second lowest among five racial groups. However, among students who are academically eligible for nonselective and 2-year institutions, the overmatch rates are 48.2% and 38.8% respectively, the highest among five racial groups, and far beyond the national average of 28.9% and 27.1%. It is the lower end on the selectively level that raises the overall overmatch rate.

Therefore, the general by-race undermatch and overmatch rate data actually disguise the college access reality that Black students face, and caution should be taken while interpreting aggregate data regarding the extent to which students undermatch. Black students are among the most underserved students in the K-12 education system, and their lower overall undermatch rate relative to the national average more so relates to their lack of eligibility for enrollment at more selective campuses. Black students who match the selectivity of their first college with their eligibility level often do so despite having been denied access to many of the resources available to their peers, and findings from this study underscore that the highest-achieving students were the most susceptible group. Considering the fact that more selective institutions are generally associated with distinctly higher graduation rates (Horn, 2006; Heil, Reisel, & Attewell, 2014) and higher earnings (Eide, Hilmer, & Showalter, 2016). outcomes, the high undermatching rates among those initially academically eligible for very selective or selective institutions are especially of concern.

Moreover, as mentioned earlier, about one-third (34.3%) of Black undergraduates are enrolled at a more selective institution than would be expected given their academic backgrounds. This finding likely connects to holistic admissions review, affirmative action policies, and diversity initiatives at campuses designed to create a more diverse student body, and it raises an interesting issue as to the extent that these campuses have the appropriate supports necessary to help students who may find themselves in over their head (or simply, overwhelmed) by the academic rigor of the curriculum. Especially in an era where affirmative action policies at some of the most elite institutions continue to be litigated (Fisher v. University of Texas, 2013; Students for Fair Admission v. Harvard, 2019), highlighting the effectiveness of such policies in providing these critical opportunities to deserving students is warranted.

Regarding SES, different from the hypothesis that students from lower SES are more likely to undermatch, findings reveal a more nuanced pattern. Students from the lowest SES quartile on average have the lowest undermatch rate (26.5%), far beyond the national average of 36.9%. Their undermatch rate among students that are academically eligible for very selective institutions is also on the lower end (65.7%, compared to national average of 61.6%). Students from the third SES quartile have the highest undermatch rate (75.4%) among students who are academically eligible for very selective institutions. Like those in the top quartile, the third quartile students are well-represented in the top three selectivity levels of eligibility, but they may not have the resources (financial) or capital (knowledge) to either afford attending the most selective campuses and/or understanding of how financial aid could help offset the costs of enrolling at more selective campuses. On the other hand, highest SES quartile students have the lowest undermatch rate, both overall rate, and at each student academic eligibility level, which suggests they have the resources as well as the cultural and social capital necessary to have both wide flexibility and critical information when deciding which college to attend.

While the undermatch rates among students from the lowest SES look optimistic, academic eligibility data uncover the hurdle that these students face. The poorest students in this sample appeared to have the least access to the most selective institutions. Among students from the lowest SES quartile, only 3.7% were academically eligible for very selective institutions, compared to 14.4% national average and 26.1% of students from the highest SES. Students from the lowest SES quartile also have the lowest percent that were academically eligible for selective and somewhat selective institutions, and the highest percent that were academically eligible for nonselective and 2-year institutions. Even after controlling for the contexts of students' high schools and neighborhoods, those from the second and third quartiles continued to have a higher likelihood of undermatching than their peers in the top SES quartile. This persistent gap undermines the notion that higher education serves as a tool for social mobility, as middle-class students who are eligible to enroll in more selective institutions decide to pursue their degrees at less prestigious campuses, which may have implications for post-college earnings.

HGLM results show that students' college choice priorities on financial aspect do influence undermatching status. Specifically, the importance on post-secondary school's low expenses and available financial aid are positively related to undermatch, that is, the more important students think low expenses and financial aid are, the more likely they would undermatch. On one hand, this confirms Belasco and Trivitte (2015)'s finding, that students, especially those from lower SES background, might be deterred by the sticker price, and worry about finances more. This also goes back to the discussion about reasons of undermatching – while some students might undermatch after rational consideration, a lot more undermatch because of inadequate information regarding how to finance college. Then as a result of undermatching, they enter colleges that might not best support their educational needs and development. On a theoretical level, this also ties to the dimension of opportunity in the Iloh (2018) model. As students consider within the landscape of higher education what options are available to them, their own financial situation limits their conceptualization of actual opportunities. They might have heard of certain "good colleges" from mass media, but a quick search about the price deterred them from seeking further information, and consequently makes them exclude this option.

Another aspect related to college priorities on finance, the importance students place on post-secondary institution job placement, is negatively related to undermatch. In other words, the more students emphasized career prospects of institutions when deciding where to enroll, the less likely they undermatched. As explained earlier, this could be due to fact that in general, higher selectivity institution graduates have higher employment rates and earnings (Rumberger & Thomas, 1993; Drydakis, 2016), and this phenomenon prompts career-focused prospective students to choose more selective institutions, which typically have stronger reputations for job placement. The implication for practice would be further discussed in the implication section.

Surprisingly, the hypothesis that the number of college information sources would significantly reduce chances of undermatch is not supported. Prior research (Baum & Schwartz, 2015) and the Iloh (2018) model proposed that students with access to multiple sources of credible information about college are more likely to make informed decisions. The result from this research shows that neither the number of personal sources (i.e., parents, siblings, other relatives, and friends) nor the number of formal sources (i.e., high school counselor, college representatives, college publications/websites, college search guides, and school or public libraries) influence the undermatching status. However, the discrepancy between this study and prior research does not indicate that college information sources do not matter, as the current study (due to the limitation of using pre-existing dataset) only considered the quantity of information source; rather, it points to the need to further focus on the quality and delivery of these college information sources. Additionally, it may be that these information sources indirectly affect undermatching, as students with access to more sources of information were also more likely to apply for financial aid or consider job placement in their search processes; the inclusion of these measures may have fully accounted for the predictive power of the college information source measures.

Lastly, the hypothesis that zip code area characteristics would influence undermatching status is partly supported. Among five demographic and socio-economic indicators, both zip code average annual income and percent of Bachelor’s degree or higher influence students’ undermatching status. Percent of bachelor’s degree or higher was once a significant factor, but its explanatory power is shared with high school characteristics. The significant influence of zip code average annual income however, still persists after controlling for all individual and high school characteristics. Firstly, this finding shows that the economic environment students live in does influence their higher education choices, no matter what their own SES background is. The mechanism, even though not clear from the current research, could be elucidated by the geography of opportunity framework. According to the Galster and Killen (1995) framework, geography decides opportunity structure through five dimensions, education, legal labor market, criminal justice system, illegal labor market, social welfare, and local social network. Youth conceptualize feasible higher education options based on the perceived high-wage legitimate employment and the social environment orientation. Therefore, this could potentially be the reason why zip code average annual income influence undermatching. Still, a more detailed model that includes more aspects of zip code characteristics, such as crime, unemployment rate differences among individuals of different educational attainment, would shed more light on this issue. In summary, this finding points to the importance of including geography in the college undermatch model, and future studies should continue to explore the way zip code area characteristics exerting influence.

Research Question Two: Choosing STEM majors

For this research question, I hypothesized that after controlling for high school and zip code area characteristics, racial background, parental SES, and students' college choice priorities on job placement, would significantly influence students' chances of choosing a STEM major. I also hypothesized that undermatched students are less likely to choose STEM majors. And lastly, I hypothesized that zip code area socioeconomic indicators would influence individual's choice of college major.

In terms of racial background, descriptive analysis shows that Asian, and Black students have higher proportion of students that chose STEM majors (32.5% and 23.0% respectively, compared to national average of 21.1%). Hispanic students have the lowest rate that chose

STEM major (15.6%). This result partly explains the different mechanisms that Black and Hispanic students are underrepresented in STEM degree holders (National Science Foundation, 2008). Black students, at the beginning of college, choose STEM majors at an even higher rate than national average; still, high attrition rates resulted in lower graduation rates in STEM. Hispanic students, while facing the same hurdle as Black students, have an additional hindrance- they also were less inclined to major in STEM at the beginning of college.

In terms of the influence of matching status on STEM major choice, undermatch was a significant positive factor that influences STEM major choice, after controlling for demographic and high school factors. It is not until individual background and zip code area characteristics enter the model that undermatching status is no longer significant. As mentioned in chapter 4, two major factors that cause this change is GPA and SAT scores. Students with higher GPA and SAT scores are more likely to undermatch, meanwhile, the model also shows that students with higher GPA and SAT scores are more likely to choose STEM. Therefore, one pressing problem for underrepresented minority students is that, students who undermatched also tend to be those with higher probability to choose STEM major. Similar to earlier discussions, as less selective institutions are typically less efficient in STEM degree production (Eagan, 2010), the high rate of choosing STEM among undermatched students signifies a potential loss of students graduating with STEM degree, especially for underrepresented minority students.

Another interesting finding regarding SES is from the HGLM result. It could be found that initially students from the lowest, second lowest, and third SES quartile are all less likely than students from the highest SES to major in STEM. However, after accounting for demographic, high school, individual academic background, and zip code area characteristics variables, students from the lower SES background no longer differ from highest SES students, while students from the third SES quartile are still significantly different. Therefore, even though students from the third SES quartile and highest SES quartile are both from a more advantaged background, their major choices patterns are quite different. Future research could look into the nuanced difference between these two SES backgrounds regarding their choices of STEM majors.

The hypothesis that the more students emphasize on post-secondary institutions' job placement records when choosing college, the more likely they would major in STEM, is supported by the HGLM results. Therefore it suggests that good career prospect of STEM majors would appeal to students, especially those who care about pragmatic aspects of educational credentials. This result could be particularly useful for initiatives to promote aspiration in STEM among high school students.

Regarding zip code area demographic and socioeconomic indicators, the hypothesis proposed that higher bachelor's degree in the area and higher annual zip income would be positively related to choosing STEM majors. As a result, HGLM analysis confirms that higher Bachelor's degree in the area does increase students' chances of choosing STEM major, while zip average annual income does not. This again points to the mechanism geography influences the choices youth make. One such mechanism is local social networks, through which the predominant group's norms and values are inculcated into the young generation (Wellman, 1972). Moreover, the informational function of local social networks might also provide educational information needed by the youth (Galster & Killen, 1995). Therefore, the more people in the district are holding a bachelor's degree of higher, the more likely they become reliable sources of information about college and major information, and provide customized and reliable suggestions for students. On the other hand, average zip annual income does not seem to influence STEM major choice, and it could be that the annual income at zip code level are not directly associated with educational attainment and familiarity with higher education system. Therefore, it could not help students with the major information they need.

Research Question Three: Employment earnings.

I hypothesized that undermatching would negatively influence students' annual earnings from employment; however, majoring in STEM will alleviate the negative influence that undermatching status has on students' annual earnings from employment.

Both hypotheses are confirmed by the HLM results. Interestingly, the initial model with only undermatching status, credential field and their interaction shows that there is almost no difference in earnings between students who undermatched and those attending a matched or more selective institution. Furthermore, individuals with a credential in STEM do earn more (21.7% more). However, after controlling for all relevant variables, things become completely different, and more importantly the interaction term between undermatching status and STEM credential is significant. The coefficients show the interesting pattern: for students not majoring in STEM, undermatched students earn significantly less (7.7% less), but for students majoring in STEM, undermatched students earn 5.9% more. Additional subgroup analysis by academic eligibility level adds more insight to this issue, and the aforementioned pattern is the most obvious among students whose academic qualifications grant them access to *very selective* institutions.

Therefore, for students who chose a STEM major, the frog pond effect seems to be obvious- being a big frog in a small pond does give an edge. Social comparison theory helps understand this phenomenon: compared to their peers that enrolled in an academically matched institution, students who enrolled in less selective institution, due to having better academic background than their classmates, might acquire higher GPA, evaluate their own academic performance better (Davis, 1966; Reitz, 1975), thus are more aspired to choose a high-academic performance career (Davis, 1966), or a more prestigious and high-income occupation (Alwin, & Otto, 1977). This mechanism might be particularly true for student choosing a STEM major, because college STEM classes not only give a lower average GPA, but also have a wider range (Rask, 2010). Moreover, previous research has found that undermatched students had more frequent interactions with faculty, and engaged in more active and collaborative learning activities (Fosnacht, 2014). In contrast, non-STEM majors typically give higher average GPA, and GPA distribution is more concentrated on the higher end, because the assignments and exams are more open to interpretation, thus there are no clearly defined correct or wrong answers (Rask, 2010). As a result, for students who could have attended a more selective institution given their academic qualifications but did not, if they major in a non-STEM field, their stronger academic background (compared to their classmates) does not make them stand out that much. Consequently, they tend to assimilate themselves into the less selective campus environment.

Moreover, the "common knowledge" that students getting a credential in STEM earn more, is only true when excluding the consideration of credential status (the credential being a bachelor's degree, associate's degree or undergraduate certificate of diploma) and academic eligibility factors. This implies that the reason why STEM credential holders earn more, is because these students are normally the ones with higher academic eligibility background (i.e., higher GPA and SAT scores), and they also tend to graduate with higher levels of credentials (getting bachelor's degree as opposed to getting associate's and undergraduate diploma). Still, as mentioned earlier, undermatching could be a good strategy for high-achieving students who want to major in STEM, potentially because their being a "big frog" in a "small pond" status ends up giving them more opportunities than they would have had in a more competitive environment. This goes back to the K-12 preparation problem, and will be discussed later in implications.

Implication for K-12 Practice and Policy

The complicated reality of undermatching and its relationship with STEM major choices revealed by this study conveys important messages regarding K-12 practice and policy. Even though the phenomenon of underrmatching is pervasive nationally and across racial groups, as well as SES groups, several measures could be taken to construct an environment that enables high school students to make better and more informed decision about attending college and choosing majors. First of all, findings suggested that compared with Asian students, underrepresented minority, especially Black and Hispanic students, are faced with a K-12 education environment that did not fully prepare them to compete for access to higher education. Low income students are facing similar problems. Therefore, K-12 education policy makers and administrators of predominantly minority schools should focus on improving academic preparation, especially college readiness of students of color, and students of lower SES background.

In addition to the prevalent problem of lack of academic preparation among underrepresented minority students, what is more disheartening is that high-achieving Black students, while overcoming environment barriers to prepare academically well, are still the most susceptible group to undermatch. It is thus of vital importance for K-12 policy makers and practitioners to provide more support to help high-achieving minority students realize and utilize the higher education opportunities that are available to them. Measures could be taken regarding the opportunities to distribute information about college and financial aid, such as having more college fairs and encourage students to attend more college fairs. Moreover, as mentioned in chapter four, high-achieving students undermatch, but those in AP courses do so less often, suggesting these courses may be providing spaces to discuss college application strategies and/or issues with financial aid. Future research might consider investigating these spaces in greater depth and determining whether there might be any strategies or best practices that could be replicated in other areas of the curriculum where more students might encounter them. Similarly, given the finding that students who engaged with the financial aid process were less likely to undermatch, high schools could incorporate something like filling out the FAFSA as part of a homework assignment. K-12 school administrators should also make effort to provide high-achieving students adequate counseling services regarding college information (McDonough, 2005). For example, reducing counselor's student loads, or ensuring that at least one counselor in every high school (or all counselors) has training related to helping students prepare for and apply to college. And when possible, connect them with outside resources, such as college access programs by federal or state department of education (Perna, 2015).

Given the finding that college choice priorities on financial aspect influence students' undermatching status, high school should explicitly address these concerns in multiple ways. Specifically, as students who emphasize on post-secondary's low expense and financial aid information are more likely to undermatch, and combining with the fact that students be deterred by "sticker price" (Perna, 2006; Rosa, 2006), lacking academic confidence (deeming themselves academically worthy of receiving financial information), and working knowledge of financial aid application (Rosa, 2006), this finding points to the need for high school counselors to familiarize themselves and students with the information about available financial aid and scholarship resources, and help students calculate the net price that is actually needed to finance college. To better support high school counselors, K-12 policy makers could also consider incorporating a common checklist, or online tracking and college resource tools throughout high school years, to track and remind students' progress toward significant college milestones, such as registering and taking SAT/ACT tests, and completing FAFSA by deadline (Phillips, Yamashiro, & Miller, 2017). States could also try implementing policy that require high school seniors to fill out the FAFSA as a graduation requirement. Currently Louisiana, Illinois and Texas are implementing such policy, and the results are promising- there is a higher high school graduation rate, and more students attend college after graduation (Leonhardt, 2019).

As findings also suggested that priority on post-secondary institutions job replacement reduces undermatching, high school counselors also ought to provide students with adequate information on the pragmatic aspect (such as employment rate, employment destination, average earnings) of possible higher education options (Hurtado, Inkelas, Briggs, & Rhee, 1997). Still, instead of putting all the burdens on counselors, responsibilities could be distributed across more teachers and staff in high school. For example, reviewing financial aid documents could be a part of English class assignments, and calculating expected family distribution to finance college could be discussed in a math class, or homeroom.

Moreover, when high school teachers and counselors are providing suggestions regarding college choice, in addition to considering students' academic background and the match between their personal background and higher education institutions, they could also take into account the potential field of major students are interested in. For students who are interested in a STEM-related major, sometimes it is not a bad idea to undermatch, as they tend to earn more than other STEM majors who matched or overmatched; but for students who are more passionate about non-STEM majors such as social sciences and humanities, counselors might need to be more careful with the suggestions they make.

Even though this research did not find direct evidence to prove that the number of college information sources would significantly reduce chances of undermatching, the importance of college information sources should not be ignored. As explained in the discussion part, it is possible that it is the quality of information sources, instead of the quantity, that exerts actual influence (Baum & Schwartz, 2015; Iloh, 2018). Considering the multiple channels of college information, both personal and formal, K-12 practitioners should not be the only actor being held accountable for reliable college information (Bryan, 2005). Instead, high school counselors could facilitate school stakeholders in implementing school-family-community partnership programs (Bryan, 2005; Hornby, 2011), and to effectively share customized information about potential higher education options and guidance on application procedures. The information dimension of the Iloh (2018) model also suggested that *who* and *how* of the message are significant. For example, a prospective student might only value recommendation from people they are closely related to, while another student only trusts official sources (such as counselors), and in this way the school-family-community partnership on information sharing would largely benefit students in helping them making a more informed decision.

Implication for Higher Education Practice and Policy

Considering the negative impact of undermatching on employment earnings, and that the undermatching situation among underrepresented minority students (especially high-achieving minorities), and low-income students is worrisome, undermatch phenomenon reveals that our current K-12 and higher education system failed to function as “the great equalizer”; instead, they facilitate perpetuating the status quo. However, findings of this research also indicated various areas where higher education policy makers and stakeholders could improve.

Current research findings indicate that applying for financial aid is related to reducing probability to undermatch and increasing the probability to choose a STEM major. Consequently, higher education policy should aim to bridge the gap between college cost and available financial resources (such as grants and scholarships, work study and subsidized loan), especially for students of lower SES background and underrepresented minority students (Stampen & Fenske, 1988). In addition to its influence on high school students' college-going decisions, financial aid has also been found to improve minority student persistence and reduces risk of dropping out (Chen & DesJardins, 2010). Another important component related to applying for financial aid is the application process. Research indicated that higher education financial system, including federal, state, institutional and private programs is confusing and complex, and often does not direct aid to students who truly need it (Spellings, 2006). Findings call for higher education policy makers to streamline financial aid application processes, and provide sufficient resource and support to help students navigate this process.

The finding that students' priority on financial aspect (low expenses, financial aid, and job placement) matters in terms of undermatching and choosing STEM majors further directed the way higher education administrator should go regarding communicating the reality of financing college. In order the improve college transparency of value, the Federal Higher Education Opportunity Act of 2008 mandated the inclusion of net price calculator for colleges that participated in federal financial aid programs. Net price calculator estimates the amount that a student needs to pay, after subtracting scholarships and grants they receives, is especially important for low-income, first-generation, and underrepresented minority students who do not have regular access to financial aid counselors (Perna, Wright-Kim, & Jiang, 2019). Still, some institutions lag behind in complying to this mandate (Cheng, 2012; Anthony, Page, & Seldin, 2016), while others are providing outdated and misleading information, such as not including full cost in the net price estimate, and lack of differentiation between grants and costs (Perna, Wright-Kim, & Jiang, 2019). Higher education administrators and policy makers should realize these problems, and make concerted effort to ensure the usability and availability of net price information. In essence, there should be greater transparency with respect to price.

The finding about geography that the zip code annual income and percentage of bachelor's degree or higher influence undermatching also provides insights on college admission practice. While it is impossible to immediately improve educational attainment and income within a region, college admission recruiters could consider putting more resources in areas with lower rates of bachelor's degree attainment so as to ensure students in those communities are learning about opportunities for aid and admission to college.

Lastly, since undermatched underrepresented minority students are more likely to choose STEM than their not-undermatched peers, it is of vital importance to make sure that their initial choice of STEM translates to final graduation with a STEM credential. Higher education policies should aim at improving STEM credential production at all level of institutions, while also pay close attention to minority students, and those came with competitive academic background. This is of especial importance to underrepresented minority students, as research has indicated that persistence in STEM is much lower for minority students (Smith, 2000). STEM departments within higher education institutions could support student by facilitating formalized mentoring programs (Foltz, Gannon, & Kirschmann, 2014), developing programs to increase faculty involvement in undergraduate training (Hurtado, Eagan, Tran, Newman,Chang,& Velasco, 2011), shifting STEM faculty pedagogy to more active learning techniques (Freeman, Eddy, &

McDonough, et al., 2014), and connecting students with outside resources such as Louis Stokes Alliances for Minority Participation (LSAMP) program.

Conclusion

Prior studies on undermatching issues have identified individual and high school factors that are related to undermatching and STEM major choice, such as racial background (Roderick, Coca, & Nagaoka, 2011; Bowen et al., 2009), SES (Smith et al., 2013; Bowen et al., 2009), academic self-efficacy (Betz & Hackett, 1983; Hackett, 1985; Lent, Brown, & Hackett, 1994), college related information source (Bowen, et al., 2009; Rodrick et al., 2008; Rodrick et al., 2009). However, few have examined the influence of broader community, along with the aforementioned factors. Moreover, the influence of postsecondary undermatching on students' outcomes, especially the different influence of undermatching in different academic field (STEM and non-STEM) also was unknown.

This study, by utilizing a more comprehensive framework that considers individual, high school and zip code level influences, contributes to the research on factors that predict undermatch, and the influence of undermatch on STEM major choice. More importantly, this research found that the influence of undermatching on students' early career earnings does differ by academic major: for students choosing a non-STEM major, attending a less selective institution probably is not a good idea; however for students that chose a STEM major, sometimes being a "big frog in small pond" might actually be beneficial economically. Still, considering the prevalent undermatching rate and low STEM rate, especially among underrepresented minority and low-income students, K-12 education and higher education stakeholders should make concerted effort to ensure that students attend higher education institutions that best fit them, and that higher education institutions provide sufficient resources for them to succeed.

Appendix A. Variables and Coding Scheme for Multilevel Models

A1. Variable coding scheme

Variable	RQ1	RQ2	RQ3	Coding Scheme
Outcome				
Undermatching status	X	X		1= yes, 0=no
Chose a STEM major 2 years after high school graduation		X		1= yes, 0=no
Employment earnings			X	Continuous: 10,000-200,000
Student demographics				
Gender: Female	X	X	X	1= yes, 0=no
Race: Asian	X	X	X	1= yes, 0=no
Race: Black	X	X	X	1= yes, 0=no
Race: Hispanic	X	X	X	1= yes, 0=no
Race: Multiracial	X	X	X	1= yes, 0=no
Race: White	X	X	X	1= yes, 0=no
SES: Lowest	X	X	X	1= yes, 0=no
SES: Second	X	X	X	1= yes, 0=no
SES: Third	X	X	X	1= yes, 0=no
SES: Highest	X	X	X	1= yes, 0=no
Native language English	X	X	X	1= yes, 0=no
Married			X	1= yes, 0=no
Academic background				
High school GPA	X	X		Continuous: 0-4
Participation in AP/IB	X	X		Continuous: : 0-18
Math self-efficacy	X	X		Continuous factor: -2.04-1.85
SAT/ACT composite score	X	X		Continuous: 420-1600
Expected level of academic achievement	X	X		1= Less than HS graduation to 8= Obtain PhD, MD or other advanced degree
College application				
Post-sec school's low expenses important to respondent	X	X		1= Not important to 3= Very important
Availability of post-sec financial aid important to respondent	X	X		1= Not important to 3= Very important
Post-sec school's job placement record important to respondent	X	X		1= Not important to 3= Very important
Whether applied for financial aid	X	X		1= yes, 0=no
College info Personal	X	X		Continuous: 0-4
College info Formal	X	X		Continuous: 0-7
Family influence				
Parents provide advice on academics	X	X		Continuous factor: -1.81-1.29
Discussion with parents: academics	X	X		Continuous factor: -2.70-1.61
Friends' influence				
# friends plan to attend 4-year college/university	X	X		Continuous: 0-5
# friends who consider grades very important	X	X		Continuous: 0-3
Residential zip code influence				
Zip % Bachelor's degree or higher	X	X	X	Continuous: 0-100
Zip % Speaks English well	X	X	X	Continuous: 0-100
Zip % Poverty	X	X	X	Continuous: 0-100

Zip % White	X	X	X	Continuous: 0-100
Zip Annual household income	X	X	X	Continuous: 22200-306400
High School characteristics				
School control (Ref: public)	X	X	X	1= yes, 0=no
School urbanicity (Ref: urban)	X	X	X	1= yes, 0=no
% minority	X	X	X	Continuous: 0-100
Student/teacher ratio	X	X	X	Continuous: 1.52-54.17
Total student enrollment	X	X	X	Continuous: 0-4533
% of graduates went to 4-year colleges	X	X	X	
% of student body is LEP or non-English proficient	X	X	X	Continuous: 0-100
School aggregated SES	X	X	X	Continuous: -0.81-1.4
Labor market characteristics				
Has a professional certification or license			X	1= yes, 0=no
Received formal employer-provided training			X	1= yes, 0=no
Formal education				
Credential field in STEM			X	1= yes, 0=no
Credential level			X	1= No credential, 2= Associate's degree or UG certificate/diploma, 3= Bachelor's degree or higher
Academically eligibility			X	1= very selective, 2-selective, 3= somewhat selective, 4= nonselective/2-year institutions
First attended institution selectivity			X	1= very selective, 2-selective, 3= somewhat selective, 4= nonselective/2-year institutions
PSI characteristics				
First PSI control (Ref: public)			X	1= yes, 0=no
First PSI sector (Ref: not-for profit)			X	1= yes, 0=no
First PSI HBCU			X	1= yes, 0=no
First PSI Total enrollment (100)			X	Continuous: 0-92281
First PSI % White (10%)			X	Continuous: 0-100
First PSI % Female (10%)			X	Continuous: 0-100
First PSI Log tuition and fees			X	Continuous: 6.25-11.04

A2. Factor Loadings and Scale Items

Factors and Items	Loading	Coding Scheme
Discussion with parents: academics (α= 0.776)		
Discuss things studied in class with parents	0.655	1= Never, 2=Sometimes, 3= Often
Discussed school courses with parents	0.651	1= Never, 2=Sometimes, 3= Often
Discussed going to college with parents	0.642	1= Never, 2=Sometimes, 3= Often
Discussed grades with parents	0.641	1= Never, 2=Sometimes, 3= Often
Discussed preparation for ACT/SAT with parents	0.626	1= Never, 2=Sometimes, 3= Often
Parents provided advice on academics (α= 0.724)		
Provide advice about plans for college entrance exams	0.851	1= Never, 2=Sometimes, 3= Often
Provide advice about applying to college/school after high school	0.684	1= Never, 2=Sometimes, 3= Often
Provide advice about selecting courses or programs	0.527	1= Never, 2=Sometimes, 3= Often

Appendix B. List of STEM majors

Biological and biomedical sciences
General biology
Biochem/biophysics/molecular biology
Botany/plant biology
Genetics
Microbiological sciences and immunology
Physiology/pathology/related sciences
Zoology/animal biology
Biological and biomedical sciences
Biological/biomedical sciences, other
Cell/cellular biology/anatomical science
Pharmacology and toxicology
Biomathematics and bioinformatics
Biotechnology
Ecology/evolution/population biology

Computer/info sciences/support technology
Computer/info technology admin/mngmnt
Computer programming
Computer science
Computer software and media applications
Computer systems analysis
Computer systems networking/telecomm
Data entry/microcomputer applications
Data processing
Information science/studies
Computer/info sciences/support, other
Computer/info sciences, general

Engineering technologies/technicians
Biomedical/medical engineering
Chemical engineering
Civil engineering
Computer engineering, general
Electrical/communications engineering
Environmental/health engineering
Mechanical engineering
Engineering, other
Engineering, general
Aerospace/aero/astronautical engineering
Agricultural/bioengineering
Architectural engineering
Ceramic sciences and engineering
Construction engineering

Engineering mechanics
Engineering physics
Engineering science
Forest engineering
Geological/geophysical engineering
Industrial engineering
Manufacturing engineering
Materials engineering
Materials science
Metallurgical engineering
Mining and mineral engineering
Naval architecture/marine engineering
Nuclear engineering
Ocean engineering
Operations research
Petroleum engineering
Polymer/plastics engineering
Surveying engineering
Systems engineering
Textile sciences and engineering

Mathematics and statistics
Mathematics
Statistics
Mathematics and statistics, other
Applied mathematics

Physical sciences
Astronomy and astrophysics
Atmospheric sciences and meteorology
Chemistry
Geological/earth sciences/geosciences
Physics
Biological and physical science
Systems science and theory
Physical sciences, other
Physical sciences, general

Science technologies/technicians
Biology/biotechnology lab technician
Nuclear/industrial radiologic technology
Physical science technology
Science technology, other
Engineering technology, general
Engineering technology, general
Architectural engineering technology

Civil engineering technology
Electrical engineering technology
Electromechanical/maintenance technology
Environmental control technology
HVAC and refrigeration technology
Industrial production technology
Quality control and safety technology
Mechanical engineering related
Mining and petroleum technology
Construction engineering technology
Engineering-related technology
Computer engineering technology
Drafting/design engineering technology
Nuclear engineering technology
Engineering-related fields
Engineering technology, other
Military technologies
Mechanical/repair technologies/techs
Health professions/clinical sciences

Appendix C: Balance Check Table for the Weighted Sample (n =3860)

		Treated	Control	%bias	bias	t	p>t
Female	Unmatched	0.56	0.51	9.90		2.73	0.006
	Matched	0.56	0.53	5.60	43.50	1.68	0.092
Asian	Unmatched	0.06	0.09	-11.40		-3.11	0.002
	Matched	0.06	0.08	-5.90	48.30	-1.83	0.068
Black	Unmatched	0.04	0.14	-34.50		-9.08	0.000
	Matched	0.04	0.10	-19.70	42.80	-6.19	0.000
Hispanic	Unmatched	0.07	0.11	-14.50		-3.92	0.000
	Matched	0.07	0.09	-8.50	41.50	-2.60	0.009
Multiracial	Unmatched	0.03	0.04	-2.00		-0.56	0.578
	Matched	0.03	0.04	-0.60	68.70	-0.19	0.846
Lowest quartile SES	Unmatched	0.12	0.15	-10.30		-2.81	0.005
	Matched	0.12	0.13	-3.80	63.20	-1.17	0.241
2nd lowest quartile SES	Unmatched	0.22	0.19	8.10		2.27	0.023
	Matched	0.22	0.20	6.40	21.90	1.93	0.053
3rd quartile SES	Unmatched	0.32	0.26	12.80		3.58	0.000
	Matched	0.32	0.28	7.50	41.60	2.26	0.024
High school GPA	Unmatched	3.38	2.89	89.40		23.86	0.000
	Matched	3.38	2.91	85.10	4.80	24.07	0.000

Appendix D. Subgroup Analysis of HLM Models Predicting Log Employment Earnings 9 Years After 10th Grade

Appendix D1. HLM Models Predicting Log Employment Earnings 9 Years After 10th Grade, Academic Eligibility-Very Selective (N=750)

	Basic			Individual background			Formal education			PSI		
	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.
Credential: STEM field	0.089	1.093		-0.041	0.960		-0.101	0.903		-0.142	0.868	
FI very selective	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
FI selective	-0.254	0.776	***	-0.204	0.815	**	-0.195	0.823	**	-0.121	0.886	
FI somewhat selective	-0.308	0.735	***	-0.211	0.810	**	-0.197	0.821	**	-0.096	0.909	
FI non-selective or 2 year	-0.150	0.861	*	-0.046	0.955		0.054	1.055		0.146	1.158	
INT: FI very selective*STEM	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
INT: FI selective*STEM	0.325	1.385	**	0.394	1.484	**	0.368	1.446	**	0.419	1.521	**
INT: FI somewhat selective*STEM	0.261	1.298	*	0.363	1.438	**	0.338	1.403	**	0.387	1.472	**
INT: FI non-selective or 2 year *STEM	0.122	1.130		0.138	1.148		0.127	1.136		0.228	1.256	
Pre-labor market influences: high school characteristics												
% of graduates went to 4-year colleges				0.052	1.054	*	0.046	1.047		0.066	1.068	**
School control (Ref: public)				0.061	1.063		0.070	1.072		0.053	1.055	
School urbanicity (Ref: urban)				0.024	1.025		0.031	1.031		0.071	1.074	
School aggregated SES				-0.085	0.918		-0.083	0.920		-0.119	0.888	
% minority (10%)				0.010	1.010		0.008	1.008		-0.002	0.998	
Student/teacher ratio (10)				0.046	1.047		0.069	1.071		0.029	1.029	
Total student enrollment (100)				-0.003	0.997		-0.003	0.997		-0.001	0.999	
% of student body is LEP or non-English proficient (10%)				0.013	1.013		0.015	1.015		0.031	1.031	
Demographics												
First Language English				-0.098	0.907		-0.093	0.911		-0.040	0.961	
Asian (Ref: White)				0.247	1.280	*	0.257	1.293	*	0.236	1.266	*
Black (Ref: White)				-0.044	0.957		-0.038	0.963		-0.222	0.801	
Hispanic (Ref: White)				-0.104	0.901		-0.107	0.899		-0.111	0.895	
Multiracial (Ref: White)				-0.010	0.990		-0.001	0.999		0.006	1.006	
Lowest quartile SES (Ref: Highest quartile)				-0.030	0.970		-0.062	0.940		-0.088	0.916	
2nd lowest quartile SES (Ref: Highest quartile)				-0.199	0.820	**	-0.204	0.816	**	-0.225	0.798	***

3rd quartile SES (Ref: Highest quartile)				-0.102	0.903		-0.085	0.919		-0.082	0.921	
Female				-0.111	0.895	*	-0.126	0.882	**	-0.109	0.896	*
Married				0.115	1.122	**	0.121	1.129	**	0.118	1.126	**
Pre-labor market influences: zip code area												
zip % Bachelor's degree or higher (10%)				0.036	1.036		0.031	1.032		0.037	1.037	
zip % Speaks english well (10%)				0.009	1.009		0.012	1.012		0.037	1.037	
zip % Poverty (10%)				0.041	1.042		0.029	1.029		0.028	1.028	
zip % White (10%)				0.002	1.002		-0.002	0.998		-0.008	0.992	
zip Log annual income				0.099	1.104		0.101	1.106		0.066	1.069	
Training												
Has a professional certification				0.060	1.062		0.076	1.079		0.082	1.085	
Has received formal employer-provided training				0.134	1.143	**	0.135	1.145	**	0.124	1.132	**
Formal education: Credential status												
UG certificate or diploma, or associate's (ref: no certificate)							-0.080	0.923		-0.063	0.939	
Bachelor's degree or higher(ref: no certificate)							0.185	1.203	***	0.176	1.192	***
Formal education: First PSI characteristics												
First PSI Total enrollment (100)										0.000	1.000	
First PSI control (Ref: public)										0.157	1.171	
First PSI sector (Ref: not-for profit)										-0.566	0.568	*
First PSI HBCU										0.065	1.067	
First PSI % White (10%)										-0.047	0.954	*
First PSI % Female (10%)										-0.030	0.970	
First PSI Log tuition and fees										-0.023	0.978	
Intercept	10.667	42917	***	10.670	43058	***	10.671	43090	***	10.665	42833	***
High school variance components (S.D)	0.0046			0.00002			0.00002			0.00003		
Level-1 Variance components (S.D)	0.2393			0.20502			0.19907			0.18699		
Deviance	1091.2			940.0			917.9			871.0		

Appendix D2. HLM Models Predicting Log Employment Earnings 9 Years After 10th Grade, Academic Eligibility- Selective (N=1060)

	Basic			Individual background			Formal education			PSI		
	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.
Credential: STEM field	0.219	1.245	**	0.193	1.213	*	0.189	1.208	*	0.189	1.208	*
FI very selective	0.098	1.103		0.058	1.059		0.061	1.063		0.030	1.030	
FI selective	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
FI somewhat selective	-0.104	0.901	*	-0.086	0.917		-0.086	0.918		-0.054	0.948	
FI non-selective or 2 year	-0.125	0.882	*	-0.086	0.918		-0.066	0.936		0.020	1.020	
INT: FI very selective*STEM	0.063	1.065		0.056	1.057		0.053	1.054		0.068	1.070	
INT: FI selective*STEM	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
INT: FI somewhat selective*STEM	0.057	1.059		-0.004	0.996		-0.004	0.996		-0.008	0.992	
INT: FI non-selective or 2 year *STEM	0.001	1.001		-0.011	0.989		0.042	1.043		0.020	1.020	
Pre-labor market influences: high school characteristics												
% of graduates went to 4-year colleges				0.010	1.010		0.008	1.008		0.004	1.004	
School control (Ref: public)				-0.049	0.952		-0.054	0.947		-0.063	0.939	
School urbanicity (Ref: urban)				0.022	1.023		0.021	1.021		0.022	1.022	
School aggregated SES				-0.071	0.932		-0.072	0.930		-0.069	0.933	
% minority (10%)				-0.026	0.974	*	-0.027	0.974	*	-0.031	0.969	**
Student/teacher ratio (10)				-0.015	0.985		-0.019	0.981		-0.019	0.982	
Total student enrollment (100)				-0.001	0.999		-0.001	0.999		-0.002	0.998	
% of student body is LEP or non-English proficient (10%)				0.031	1.032		0.028	1.028		0.030	1.031	
Demographics												
First Language English				-0.084	0.920		-0.081	0.922		-0.078	0.925	
Asian (Ref: White)				0.015	1.016		0.027	1.027		0.000	1.000	
Black (Ref: White)				0.085	1.089		0.085	1.089		0.063	1.065	
Hispanic (Ref: White)				0.052	1.053		0.051	1.052		0.042	1.042	
Multiracial (Ref: White)				-0.073	0.929		-0.071	0.932		-0.081	0.922	
Lowest quartile SES (Ref: Highest quartile)				-0.189	0.828	**	-0.191	0.826	**	-0.179	0.836	**
2nd lowest quartile SES (Ref: Highest quartile)				-0.099	0.906		-0.097	0.907		-0.092	0.912	
3rd quartile SES (Ref: Highest quartile)				-0.106	0.899	**	-0.106	0.899	**	-0.105	0.900	**
Female				-0.155	0.856	***	-0.150	0.861	***	-0.145	0.865	***

Married				0.108	1.114	**	0.110	1.116	***	0.119	1.126	***
Pre-labor market influences: zip code area												
zip % Bachelor's degree or higher (10%)				0.017	1.018		0.019	1.020		0.012	1.012	
zip % Speaks english well (10%)				-0.007	0.993		-0.008	0.992		0.001	1.001	
zip % Poverty (10%)				0.033	1.034		0.033	1.034		0.030	1.030	
zip % White (10%)				-0.009	0.991		-0.009	0.991		-0.013	0.987	
zip Log annual income				0.185	1.203	*	0.186	1.205	*	0.173	1.189	*
Training												
Has a professional certification				0.104	1.110	**	0.102	1.107	**	0.105	1.110	**
Has received formal employer-provided training				0.057	1.059		0.055	1.057		0.070	1.073	*
Formal education: Credential status												
UG certificate or diploma, or associate's (ref: no certificate)							-0.112	0.894		-0.090	0.914	
Bachelor's degree or higher(ref: no certificate)							0.010	1.011		0.002	1.002	
Formal education: First PSI characteristics												
First PSI Total enrollment (100)										0.000	1.000	
First PSI control (Ref: public)										-0.022	0.978	
First PSI sector (Ref: not-for profit)										-0.159	0.853	
First PSI HBCU										0.129	1.138	
First PSI % White (10%)										-0.019	0.981	
First PSI % Female (10%)										-0.011	0.989	
First PSI Log tuition and fees										0.074	1.077	
Intercept	10.530	37407	***	10.534	37570	***	10.534	37585	***	10.532	37508	***
High school variance components (S.D)	0.0001			0.00005			0.00005			0.00004		
Level-1 Variance components (S.D)	0.183			0.15786			0.15727			0.15421		
Deviance	1210.4			1055.8			1051.8			1027.9		

Appendix D3. HLM Models Predicting Log Employment Earnings 9 Years After 10th Grade, Academic Eligibility- Somewhat Selective (N=950)

	Basic			Individual background			Formal education			PSI		
	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.
Credential: STEM field	0.152	1.164		0.025	1.025		-0.014	0.986		0.012	1.012	
FI very selective	-0.012	0.988		-0.043	0.958		-0.019	0.981		-0.061	0.940	
FI selective	0.161	1.175	*	0.113	1.120		0.105	1.111		0.079	1.083	
FI somewhat selective	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
FI non-selective or 2 year	-0.111	0.895	*	-0.090	0.914	*	-0.053	0.948		-0.066	0.936	
INT: FI very selective*STEM	0.056	1.057		0.236	1.266		0.186	1.205		0.182	1.199	
INT: FI selective*STEM	-0.070	0.933		0.045	1.046		0.034	1.034		-0.013	0.987	
INT: FI somewhat selective*STEM	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
INT: FI non-selective or 2 year *STEM	0.114	1.121		0.213	1.238		0.257	1.293		0.261	1.298	
Pre-labor market influences: high school characteristics												
% of graduates went to 4-year colleges				0.011	1.011		0.011	1.011		0.008	1.008	
School control (Ref: public)				0.047	1.048		0.030	1.031		0.034	1.035	
School urbanicity (Ref: urban)				-0.087	0.917	*	-0.079	0.924		-0.068	0.934	
School aggregated SES				0.018	1.018		0.016	1.017		0.023	1.023	
% minority (10%)				-0.013	0.987		-0.017	0.983		-0.016	0.984	
Student/teacher ratio (10)				-0.003	0.997		-0.015	0.985		-0.027	0.973	
Total student enrollment (100)				0.004	1.004		0.003	1.003		0.004	1.004	
% of student body is LEP or non-English proficient (10%)				0.028	1.028		0.032	1.033		0.029	1.029	
Demographics												
First Language English				-0.215	0.807	**	-0.218	0.805	**	-0.210	0.811	**
Asian (Ref: White)				-0.052	0.949		-0.045	0.956		-0.040	0.961	
Black (Ref: White)				-0.033	0.968		-0.027	0.973		0.045	1.046	
Hispanic (Ref: White)				-0.073	0.930		-0.082	0.921		-0.075	0.928	
Multiracial (Ref: White)				-0.173	0.841	*	-0.163	0.850		-0.153	0.858	
Lowest quartile SES (Ref: Highest quartile)				-0.025	0.975		-0.018	0.982		-0.021	0.979	
2nd lowest quartile SES (Ref: Highest quartile)				-0.012	0.988		-0.015	0.986		-0.024	0.976	
3rd quartile SES (Ref: Highest quartile)				0.091	1.096	*	0.087	1.091	*	0.088	1.092	*
Female				-0.117	0.889	**	-0.116	0.890	**	-0.105	0.900	**

Married				0.164	1.178	***	0.163	1.177	***	0.163	1.177	***
Pre-labor market influences: zip code area												
zip % Bachelor's degree or higher (10%)				-0.012	0.988		-0.015	0.985		-0.019	0.981	
zip % Speaks english well (10%)				-0.036	0.965		-0.044	0.957		-0.035	0.966	
zip % Poverty (10%)				0.005	1.005		0.003	1.003		0.008	1.008	
zip % White (10%)				-0.003	0.997		-0.004	0.996		-0.007	0.993	
zip Log annual income				0.099	1.104		0.091	1.096		0.101	1.106	
Training												
Has a professional certification				0.127	1.135	**	0.130	1.139	**	0.126	1.134	**
Has received formal employer-provided training				0.100	1.106	**	0.098	1.103	**	0.092	1.097	**
Formal education: Credential status												
UG certificate or diploma, or associate's (ref: no certificate)							-0.065	0.937		-0.087	0.916	
Bachelor's degree or higher(ref: no certificate)							0.100	1.106	*	0.082	1.085	
Formal education: First PSI characteristics												
First PSI Total enrollment (100)										0.000	1.000	
First PSI control (Ref: public)										0.015	1.015	
First PSI sector (Ref: not-for profit)										-0.133	0.875	
First PSI HBCU										-0.386	0.680	***
First PSI % White (10%)										-0.001	0.999	
First PSI % Female (10%)										-0.062	0.940	**
First PSI Log tuition and fees										-0.012	0.988	
Intercept	10.393	32639	***	10.402	32939	***	10.403	32943	***	10.404	32990	***
High school variance components (S.D)	0.01713			0.00019			0.0004			0.0037		
Level-1 Variance components (S.D)	0.18857			0.18638			0.184			0.1717		
Deviance	1205.7			1100.7			1090.0			1048.0		

Appendix D4. HLM Models Predicting Log Employment Earnings 9 Years After 10th Grade, Academic Eligibility- Nonselective or two-year institution (N=1100)

	Basic			Individual background			Formal education			PSI		
	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.	Coef.	exp(b)	sig.
Credential: STEM field	0.100	1.106		0.038	1.039		-0.010	0.990		-0.004	0.996	
FI very selective	0.141	1.151		0.075	1.078		0.044	1.045		0.034	1.035	
FI selective	0.002	1.002		0.015	1.015		-0.005	0.995		-0.033	0.967	
FI somewhat selective	0.053	1.054		0.080	1.083		0.056	1.057		0.033	1.034	
FI non-selective or 2 year	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
INT: FI very selective*STEM	-0.316	0.729		-0.253	0.777		-0.267	0.766		-0.274	0.760	
INT: FI selective*STEM	0.073	1.075		0.084	1.087		0.055	1.056		0.050	1.051	
INT: FI somewhat selective*STEM	0.256	1.292	*	0.327	1.386	**	0.312	1.366	**	0.313	1.368	**
INT: FI non-selective or 2 year *STEM	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)	(REF)
Pre-labor market influences: high school characteristics												
% of graduates went to 4-year colleges				-0.005	0.995		-0.006	0.994		-0.007	0.993	
School control (Ref: public)				0.120	1.128		0.112	1.119		0.095	1.099	
School urbanicity (Ref: urban)				0.016	1.017		0.012	1.012		0.005	1.005	
School aggregated SES				-0.247	0.781	**	-0.248	0.780	**	-0.241	0.786	**
% minority (10%)				0.000	1.000		0.001	1.001		0.000	1.000	
Student/teacher ratio (10)				-0.088	0.916		-0.086	0.918		-0.077	0.926	
Total student enrollment (100)				0.008	1.008	**	0.007	1.007	**	0.007	1.007	**
% of student body is LEP or non-English proficient (10%)				-0.012	0.988		-0.012	0.988		-0.009	0.991	
Demographics												
First Language English				-0.193	0.824	*	-0.195	0.823	*	-0.190	0.827	*
Asian (Ref: White)				-0.198	0.820		-0.201	0.818		-0.193	0.825	
Black (Ref: White)				-0.121	0.886		-0.125	0.883	*	-0.113	0.893	
Hispanic (Ref: White)				-0.233	0.792	**	-0.238	0.788	**	-0.238	0.788	**
Multiracial (Ref: White)				-0.087	0.917		-0.084	0.920		-0.095	0.909	
Lowest quartile SES (Ref: Highest quartile)				-0.126	0.881	*	-0.124	0.884	*	-0.125	0.883	*
2nd lowest quartile SES (Ref: Highest quartile)				-0.018	0.982		-0.016	0.984		-0.016	0.984	
3rd quartile SES (Ref: Highest quartile)				-0.070	0.933		-0.069	0.934		-0.072	0.931	

Female				-0.256	0.774	***	-0.257	0.774	***	-0.268	0.765	***
Married				0.197	1.218	***	0.198	1.219	***	0.202	1.224	***
Pre-labor market influences: zip code area												
zip % Bachelor's degree or higher (10%)				-0.019	0.981		-0.018	0.982		-0.025	0.976	
zip % Speaks english well (10%)				-0.023	0.977		-0.022	0.979		-0.020	0.980	
zip % Poverty (10%)				0.011	1.011		0.010	1.010		0.007	1.007	
zip % White (10%)				-0.010	0.991		-0.010	0.990		-0.012	0.988	
zip Log annual income				0.215	1.240	**	0.214	1.238	**	0.218	1.243	**
Training												
Has a professional certification				0.078	1.081	*	0.081	1.084	*	0.083	1.086	*
Has received formal employer-provided training				0.066	1.068	*	0.062	1.064		0.059	1.061	
Formal education: Credential status												
UG certificate or diploma, or associate's (ref: no certificate)							0.054	1.056		0.055	1.057	
Bachelor's degree or higher(ref: no certificate)							0.111	1.117		0.122	1.129	
Formal education: First PSI characteristics												
First PSI Total enrollment (100)										0.000	1.000	
First PSI control (Ref: public)										0.000	1.000	
First PSI sector (Ref: not-for profit)										-0.068	0.935	
First PSI HBCU										-0.144	0.866	
First PSI % White (10%)										-0.005	0.995	
First PSI % Female (10%)										0.026	1.027	
First PSI Log tuition and fees										0.026	1.026	
Intercept	10.293	29528	***	10.287	29362	***	10.287	29335	***	10.286	29327	***
High school variance components (S.D)	0.02787			0.0058			0.00276			0.00031		
Level-1 Variance components (S.D)	0.20522			0.1896			0.19345			0.19608		
Deviance	1518.9			1326.7			1327.4			1324.0		

Reference

Acemoglu, D., & Autor, D. (2011). *Lectures in labor economics*. Retrieved from https://economics.mit.edu/files/4689

Adelman, C. (1999). Answers in the Tool Box. Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment. Washington, D.C.: U.S. Department of Education

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: US Department of Education.

Agnello, R. J., & Hunt Jr, J. W. (1976). The Impact of a Part-Time Graduate Degree and Early-Career Earnings on Late-Career Earnings. *Journal of Human Resources*, 209-218.

Allensworth, E. M. (2006). *From high school to the future: A first look at Chicago Public School graduates' college enrollment, college preparation, and graduation from four-year colleges*. Chicago, IL: Consortium on Chicago School Research.

Alicke, M. D., Zell, E., & Bloom, D. L. (2010). Mere categorization and the frog-pond effect. *Psychological Science*, *21*(2), 174-177.

Alwin, D. F., & Otto, L. B. (1977). High school context effects on aspirations. *Sociology of Education*, 259-273.

Anderson, E., & Kim, D. (2006). *Increasing the success of minority students in science and technology*. Washington, DC: American Council on Education.

Anthony, A. M., Page, L. C., & Seldin, A. (2016). In the Right Ballpark? Assessing the Accuracy of Net Price Calculators. *Journal of Student Financial Aid*, *46*(2).

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, *166*(1), 3-16.

Aries, E., & Seider, M. (2005). The interactive relationship between class identity and the college experience: The case of lower income students. *Qualitative Sociology*, *28*(4), 419-443.

Astin, A. W. (1993). *What matters in college?: Four critical years revisited* (Vol. 1). San Francisco, CA: Jossey-Bass.

Aud, S., KewalRamani, A., & Frolich, L. (2011). *America's youth: Transitions into adulthood (NCES 2012-026).* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Avery, C., & Kane, T. J. (2004). Student perceptions of college opportunities. The Boston COACH program. In *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 355-394). Chicago, IL: University of Chicago Press.

Avery, C. (2010). *The effects of college counseling on high-achieving, low-income students* (No. w16359). New York, NY: National Bureau of Economic Research.

Backes, B. (2012). Do affirmative action bans lower minority college enrollment and attainment? Evidence from statewide bans. *Journal of Human Resources*, *47*(2), 435-455.

Bailey, Thomas, and Melinda Mechur Karp (2003). *Promoting College Access and Success: A Review of Credit-Based Transition Programs*. Washington, D.C: U.S. Department of Education, Office of Adult and Vocational Education

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Barron, J. M., Black, D. A., & Loewenstein, M. A. (1987). Employer size: The implications for search, training, capital investment, starting wages, and wage growth. *Journal of Labor Economics*, *5*(1), 76-89.

Barron's Educational Series (2006). *Barron's profiles of american Colleges*. Hauppauge, New York: Barron's Educational Series.

Bastedo, M. N., & Bowman, N. A. (2010). US News & World Report college rankings: Modeling institutional effects on organizational reputation. *American Journal of Education*, *116*(2), 163-183.

Bastedo, M. N., & Jaquette, O. (2011). Running in place: Low-income students and the dynamics of higher education stratification. *Educational Evaluation and Policy Analysis*, *33*(3), 318-339.

Bastedo, M. N., & Flaster, A. (2014). Conceptual and methodological problems in research on college undermatch. *Educational Researcher*, *43*(2), 93-99.

Baum, S., Ma, J., & Payea, K. (2010). Education pays: The benefits of education for individuals and society. *Trends in higher education series*. New York: The College Board.

Baum, S., & Schwartz, S. (2015). Student aid, student behavior, and educational attainment. In B. L. Castleman, S. Baum, & S. Schwartz (Eds.), *Decision making for student success: Behavioral insights to improve college access and persistence* (pp. 38–62). New York: Routledge.

Becker, G. S. (1975). *Human Capital: A theoretical and empirical analysis, with special reference to education*. New York, NY: National Bureau of Economic Research.

Becker, G. S. (1994). Human capital revisited. In *Human Capital: A theoretical and empirical analysis with special reference to education (3rd Edition)* (pp. 15-28). Chicago, IL: The university of Chicago press.

Beede, D. N., Julian, T. A., Khan, B., Lehrman, R., McKittrick, G., Langdon, D., & Doms, M. E. (2011). Education supports racial and ethnic equality in STEM. *Economics and Statistics Administration Issue Brief*, (05-11).

Beffy, M., Fougere, D., & Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter?. *Review of Economics and Statistics*, *94*(1), 334-347.

Belasco, A. S., & Trivette, M. J. (2015). Aiming low: Estimating the scope and predictors of postsecondary undermatch. *The Journal of Higher Education*, *86*(2), 233-263.

Berryman, S. E. (1983). *Who will do science? Trends, and their causes in minority and female representation among holders of advanced degrees in science and mathematics.* New York, NY: Rockefeller Foundation.

Bettis, P. J. (1996). Urban students, liminality, and the postindustrial context. *Sociology of Education*, 105-125.

Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational behavior*, *23*(3), 329-345.

Black, D. A., & Smith, J. A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of econometrics*, *121*(1-2), 99-124.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *Bmj*, *314*(7080), 572.

Bottia, M. C., Stearns, E., Mickelson, R. A., Moller, S., & Valentino, L. (2015). Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM. *Economics of Education Review*, *45*, 14-27.

Bound, J., Lovenheim, M. F., & Turner, S. (2012). Increasing time to baccalaureate degree in the United States. *Education*, *7*(4), 375-424.

Bourdieu, P. (1986). The forms of capital. In J.G. Richardson (ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). New York: Greenwood Press.

Bowen, W., Kurzweil, M., & Tobin, E. (2005). From bastions of privilege to engines of opportunity. *Chronicle of Higher Education*, *51*(25), B18-B18.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.

Bowers, T. A., & Pugh, R. C. (1973). Factors underlying college choice by students and parents. *Journal of College Student Personnel13*(3), 220-224.

Bowman, N. A. & Bastedo, M. N., (2016, April). What role do admissions office diversity and practices play in equitable decisions? Paper presented at the American Educational Research Association, Washinton, DC.

Bronfenbrenner, U., & Morris, P. (1998). The ecology of developmental processes. In *Theorectical models of human development*. Wiley.

Bryan, J. (2005). Fostering educational resilience and achievement in urban schools through school-family-community partnerships. *Professional School Counseling*, 219-227.

Bryan, J., Moore-Thomas, C., Day-Vines, N. L., & Holcomb-McCoy, C. (2011). School counselors as social capital: The effects of high school college counseling on college application rates. *Journal of Counseling & Development*, *89*(2), 190-199.

Brewer, D. J., Eide, E. R., & Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-Cohort Evidence on the Effects of College Type on Earnings. *The Journal of Human Resources*, *34*(1), 104–123. doi:10.2307/146304

Bureau of Labor Statistics, 2014. *College enrollment and work activity of 2013 High School Graduates.* Retrieved March 3, 2015, from http://www.bls.gov/news.release/hsgec.nr0.htm

Cabrera, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process. *New Directions for Institutional Research*, *27*(3), 5-22.

Cabrera, A. F., & La Nasa, S. M. (2001). On the path to college: Three critical tasks facing America's disadvantaged. *Research in Higher Education*,*42*(2), 119-149.

Card, D., & Krueger, A. B. (1992). School quality and black-white relative earnings: A direct assessment. *The Quarterly Journal of Economics*, *107*(1), 151-200.

Carlsson, M., & Rooth, D. O. (2012). Revealing taste-based discrimination in hiring: a correspondence testing experiment with geographic variation. *Applied Economics Letters*, *19*(18), 1861-1864.

Carnevale, A. P., & Rose, S. J. (2004). Socioeconomic status, race/ethnicity and selective college admissions. In R. D. Kahlenberg (Ed.), *America's untapped resource: Low income students in higher education* (pp. 101–156). Washington, DC: Century Foundation Press

Carnevale, A. P., Cheah, B., & Strohl, J. (2012). *Hard times: College majors, unemployment and earnings.* Washington, DC: Georgetown Public Policy Institute.

Cellini, S. R., & Turner, N. (2019). Gainfully employed? Assessing the employment and earnings of for-profit college students using administrative data. *Journal of Human Resources*, *54*(2), 342-370.

Chapman, D. W. (1981). A model of student college choice. *The Journal of Higher Education*, *52*(5), 490-505.

Charles, K., & Guryan, J. (2009). Taste-based discrimination. In *The New Palgrave Dictionary of Economics*. Basingstoke, United Kingdom: Palgrave Macmillan.

Chay, K. Y., & Lee, D. S. (2000). Changes in relative wages in the 1980s: Returns to observed and unobserved skills and black–white wage differentials. *Journal of Econometrics*, *99*(1), 1-38.

Chen, X. (2009). Students Who Study Science, Technology, Engineering, and Mathematics (STEM) in Postsecondary Education. Stats in Brief. NCES 2009-161. *National Center for Education Statistics*.

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001. *National Center for Education Statistics*.

Cheng, D. (2012). Adding It All up 2012: Are College Net Price Calculators Easy to Find, Use, and Compare?. *Institute for College Access & Success*.

Choy, S. (2001). *Students whose parents did not go to college: Postsecondary access, persistence, and attainment*. Washington, DC: National Center for Education Statistics.

Cornwell, C., & Rupert, P. (1997). Unobservable individual effects, marriage and the earnings of young men. *Economic Inquiry*, *35*(2), 285-294.

Dale, S., & Krueger, A. B. (2011). *Estimating the return to college selectivity over the career using administrative earnings data*. New York, NY: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17159

Costigan, A. T., & Grey, L. (Eds.). (2014). *Demythologizing educational reforms: Responses to the political and corporate takeover of education*. New York: Routledge.

Dale, S., & Krueger, A. B. (2011). *Estimating the return to college selectivity over the career using administrative earnings data* (No. w17159). National Bureau of Economic Research.

Darity Jr, W. A. (1982). The human capital approach to black-white earnings inequality: Some unsettled questions. *Journal of Human Resources*, 72-93.

Darling, D. (2005). Improving minority student achievement by making cultural connections. *Middle School Journal, 36*(5), 46-50.

Davison, M. L., Jew, G. B., & Davenport Jr, E. C. (2014). Patterns of SAT scores, choice of STEM major, and gender. *Measurement and Evaluation in Counseling and Development*, *47*(2), 118-126.

Daymont, T. N., & Andrisani, P. J. (1984). Job preferences, college major, and the gender gap in earnings. *Journal of Human Resources*, 408-428.

de los Santos Jr, A., & Wright, I. (1990). Maricopa's Swirling Students: Earning One-Third of Arizona State's Bachelor's Degrees. *Community, Technical, and Junior College Journal, 60*(6), 32-34.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.

Ding, W., & Lehrer, S. F. (2014). Understanding the role of time-varying unobserved ability heterogeneity in education production. *Economics of Education Review*, *40*, 55-75.

Dowd, A. C., & Bensimon, E. M. (2015). *Engaging the" race Question": Accountability and Equity in US Higher Education*. New York, NY: Teachers College Press.

Drake, T. A. (2014). The effect of community linguistic isolation on language-minority student achievement in high school. *Educational Researcher*, *43*(7), 327-340.

Drydakis, N. (2016). The effect of university attended on graduates' labour market prospects: A field study of Great Britain. *Economics of Education Review, 52*, 192-208.

Dubow, E. F., Boxer, P., & Huesmann, L. R. (2009). Long-term effects of parents' education on children's educational and occupational success: Mediation by family interactions, child aggression, and teenage aspirations. *Merrill-Palmer quarterly*, *55*(3), 224.

Dyce, C. M., Albold, C., & Long, D. (2012). Moving from college aspiration to attainment: Learning from one college access program. *The High School Journal*, 152-165.

Eagan Jr, M. K. (2010). *Moving beyond Frontiers: How Institutional Context Affects Degree Production and Student Aspirations in STEM*. ProQuest LLC. 789 East Eisenhower Parkway, PO Box 1346, Ann Arbor, MI 48106.

Eagan Jr, M. K., Hurtado, S., Chang, M. J., Garcia, G. A., Herrera, F. A., & Garibay, J. C. (2013). Making a difference in science education: the impact of undergraduate research programs. *American educational research journal*, *50*(4), 683-713.

Eagan, K., Hurtado, S., & Chang, M. (2010, October). What matters in STEM: Institutional contexts that influence STEM bachelor's degree completion rates. In *annual meeting of the Association for the Study of Higher Education, Indianapolis, IN*.

Eagan, K., Stolzenberg, E,, Ramirez, J., Aragon, M., Suchard, M., & Rios-Aguilar, C. (2016). *The American freshman: Fifty-year trends, 1966-2015*. Higher Education Research Institute, Graduate School of Education & Information Studies, University of California, Los Angeles.

Eide, E. R., Hilmer, M. J., & Showalter, M. H. (2016). Is it where you go or what you study? The relative influence of college selectivity and college major on earnings. *Contemporary Economic Policy*, *34*(1), 37-46.

Ehrenberg, R. G. (2010). Analyzing the factors that influence persistence rates in STEM field, majors: Introduction to the symposium. *Economics of Education Review*, *29*(6), 888-891.

Eide, E. R., Hilmer, M. J., & Showalter, M. H. (2016). Is it where you go or what you study? The relative influence of college selectivity and college major on earnings. *Contemporary Economic Policy*, *34*(1), 37-46.

Elliott, R., Strenta, A. C., Adair, R., Matier, M., & Scott, J. (1996). The role of ethnicity in choosing and leaving science in highly selective institutions. *Research in Higher Education*, *37*(6), 681-709.

Ellwood, D., & Kane, T. J. (2000). Who is getting a college education? Family background and the growing gaps in enrollment. *Securing the future: Investing in children from birth to college*, 283-324. Engberg, M. E., & Wolniak, G. C. (2010). Examining the effects of high school contexts on postsecondary enrollment. *Research in Higher Education*, *51*(2), 132-153.

Epstein, J. L., & Sanders, M. G. (2000). Connecting home, school, and community. In *Handbook of the sociology of education* (pp. 285-306). Springer, Boston, MA.

Espenshade, T. J., Chung, C. Y., & Walling, J. L. (2004). Admission preferences for minority students, athletes, and legacies at elite universities. *Social Science Quarterly*, *85*(5), 1422-1446.

Feldman, K. A., Smart, J. C., & Ethington, C. A. (1999). Major field and person-environment fit: Using Holland's theory to study change and stability of college students. *The Journal of Higher Education*, *70*(6), 642-669.

Fann, A., McClafferty Jarsky, K., & McDonough, P. M. (2009). Parent involvement in the college planning process: A case study of P-20 collaboration. *Journal of Hispanic Higher Education*, *8*(4), 374-393.

Ferber, M. A., & Spaeth, J. L. (1984). Work characteristics and the male-female earnings gap. *The American Economic Review*, *74*(2), 260-264.

Fetter, J. H. (1997). *Questions and admissions: Reflections on 100,000 admissions decisions at Stanford*. Palo Alto, CA: Stanford University Press.

Fisher v. University of Texas, 570 U.S. 297 (2013)

Fitzgerald, B. (2004). Federal financial aid and college access. *Readings on equal education*, *19*, 1-28.

Foltz, L. G., Gannon, S., & Kirschmann, S. L. (2014). Factors that contribute to the persistence of minority students in STEM Fields. *Planning for Higher Education*, *42*(4), 1-13.

Fosnacht, K. (2015). Undermatching and the first-year experience: examining effect heterogeneity, presented at the annual meeting of the Association for the Study of Higher Education, Denver, CO, 2015.

Fox, M. (1993). Is it a good investment to attend an elite private college?. *Economics of Education Review*, *12*(2), 137-151.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

Fry, R. (2007). *How far behind in math and reading are English language learners?* Washington, DC: Pew Hispanic Center.

Galster, G. C., & Killen, S. P. (1995). The geography of metropolitan opportunity: A reconnaissance and conceptual framework. *Housing Policy Debate, 6*(1), 7-43.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality?. *American journal of Sociology*, 1146-1183.

Gansemer-Topf, A. M., & Schuh, J. H. (2006). Institutional selectivity and institutional expenditures: Examining organizational factors that contribute to retention and graduation. *Research in Higher Education, 47*(6), 613-642.

Goldthorpe, J., & Jackson, M. (2008). Education-based meritocracy: The barriers to its realization. In Lareau, A., & Conley, D. (2008). *Social Class: How Does It Work?* New York, NY: Russell Sage Foundation.

Goldstein, H., Browne, W., & Rasbash, J. (2002). Partitioning variation in multilevel models. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *1*(4), 223-231.

Gonzalez, K. P., Stoner, C., & Jovel, J. E. (2003). Examining the role of social capital in access to college for Latinas: Toward a college opportunity framework. *Journal of Hispanic Higher Education*, *2*(2), 146-170.

Goodman, J., Hurwitz, M., & Smith, J. (2014). Initial college choice and degree completion rates: Using admissions test score thresholds to estimate undermatch penalties. Retrieved from http://www.sole-jole.org/14064.pdf

Gose, B. (1995). Double dippers'. *Chronicle of Higher Education*, *41*(47), A27-A27.

Goyette, K. A., & Mullen, A. L. (2006). Who studies the arts and sciences? Social background and the choice and consequences of undergraduate field of study. *The Journal of Higher Education*, *77*(3), 497-538.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, *31*(2), 197-218.

Granfield, R. (1991). Making it by faking it: Working-class students in an elite academic environment. *Journal of Contemporary Ethnography*, *20*(3), 331-351.

Greene, J. P., & Forster, G. (2003). *Public high school graduation and college readiness rates in the United States. Education Working Paper No. 3*. Atlanta, GA: *Center for Civic Innovation*.

Green, D. A., & Riddell, W. C. (2003). Literacy and earnings: an investigation of the interaction of cognitive and unobserved skills in earnings generation. *Labour Economics*, *10*(2), 165-184.

Greenwood, G. E., & Hickman, C. W. (1991). Research and practice in parent involvement: Implications for teacher education. *The elementary school journal*, *91*(3), 279-288.

Gunderson, M., & Oreopoulos, P. (2010). Returns to education in developed countries. *Economics of education*, 37-43.

Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of counseling psychology*, *32*(1), 47.

Haider, S., & Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American Economic Review*, *96*(4), 1308-1320.

Hallinan, M. T. (1992). The organization of students for instruction in the middle school. *Sociology of Education*, 114-127.

Hanushek, E. A., & Woessmann, L. (2016). Knowledge capital, growth, and the East Asian miracle. *Science*, *351*(6271), 344-345.

Haveman, R. H., & Smeeding, T. M. (2006). The role of higher education in social mobility. *The Future of children*, *16*(2), 125-150.

Hecker, D. E. (1995). Earnings of college graduates, 1993. *Monthly Lab. Rev.*, *118*, 3.

Hearn, J. C. (1991). Academic and nonacademic influences on the college destinations of 1980 high school graduates. *Sociology of education*, 158-171.

Heil, S., Reisel, L., & Attewell, P. (2014). College selectivity and degree completion. *American Educational Research Journal*, *51*(5), 913-935.

Higher Education Research Institute (2010). *Degrees of success: Bachelor's degree completion rates among initial STEM majors.* Retrieved on Dec 14, 2019, from https://heri.ucla.edu/nih/downloads/2010-Degrees-of-Success.pdf

Hill, C. (2015). *The simple truth about the gender pay gap*. Washington, DC: American Association of University Women.

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics*, *94*(3), 712-722.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *Review of Economics and Statistics, 91*(4), 717–724. doi:10.1162/rest.91.4.717

Horn, C. L., Flores, S. M., & Orfield, G. (2003). P*ercent plans in college admissions: A comparative analysis of three states' experiences*. Cambridge, MA: Civil Rights Project, Harvard University.

Horn, L. (2006). *Placing college graduation rates in context: How 4-year college graduation rates vary with selectivity and the size of low-income enrollment. Postsecondary education descriptive analysis report*. NCES 2007-161. Washington, DC: National Center for Education Statistics.

Hornby, G. (2011). *Parental involvement in childhood education: Building effective school-family partnerships*. Springer Science & Business Media.

Hossler, D., and Gallagher, L. (1987). Studying college choice: A three-phase model and the implication for policy makers. *College and University, 2*:201–221.

Hossler, D., Schmit, J., and Vesper, N. (1999). *Going to college. How social, economic, and educational factors influence the decisions students make*. Baltimore, MD: Johns Hopkins University Press.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Hoxby, C. M. (2009). *The changing selectivity of American colleges* (No. w15446). National Bureau of Economic Research.

Hoxby, C. M., & Avery, C. (2012). *The missing" one-offs": The hidden supply of high-achieving, low income students* (No. w18586). New York, NY: National Bureau of Economic Research.

Hoxby, C., & Turner, S. (2013). *Expanding college opportunities for high-achieving, low income students*. Stanford, CA: Stanford Institute for Economic Policy Research.

Hoxby, C., & Avery, C. (2013). *The missing" one-offs": The hidden supply of high-achieving, low-income students*. New York, NY: National Bureau of Economic Research.

Hu, S. (2003). Educational aspirations and postsecondary access and choice. *education policy analysis archives*, *11*, 14.

Hughes, R. P. (2013). *Essays on admissions matching and associated outcomes in the market for higher education in the united states* (Order No. 3576462). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (1467527682). Retrieved from https://search.proquest.com/docview/1467527682?accountid=14512

Hurtado, S., Inkelas, K. K., Briggs, C., & Rhee, B. S. (1997). Differences in college access and choice among racial/ethnic groups: Identifying continuing barriers. *Research in Higher Education*, *38*(1), 43-75.

Hurtado, S., Eagan, M. K., Tran, M. C., Newman, C. B., Chang, M. J., & Velasco, P. (2011). “We do science here”: Underrepresented students’ interactions with faculty in different college contexts. *Journal of social issues*, *67*(3), 553-579.

Iloh, C. (2018). Toward a new model of college “choice” for a twenty-first-century context. *Harvard Educational Review*, *88*(2), 227-244.

Iloh, C. (2019). An Alternative to College" Choice" Models and Frameworks: The Iloh Model of College-Going Decisions and Trajectories. *College and University*, *94*(4), 2-9.

Jacobs, J. A. (1995). Gender and academic specialties: Trends among recipients of college degrees in the 1980s. *Sociology of Education*, 81-98.

Jacobs, J. A. (1986). The sex-segregation of fields of study: Trends during the college years. *The Journal of Higher Education*, *57*(2), 134-154.

James, E., Alsalam, N., Conaty, J. C., & To, D. L. (1989). College quality and future earnings: Where should you send your child to college?. *The American Economic Review*, *79*(2), 247-252.

Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement: A meta-analysis. *Urban education*, *42*(1), 82-110.

John, E. P. S., Hu, S., & Weber, J. (2001). State policy and the affordability of public higher education: The influence of state grants on persistence in Indiana. *Research in Higher Education*, *42*(4), 401-428.

John, E. P. S., & Asker, E. H. (2003). *Refinancing the college dream: Access, equal opportunity, and justice for taxpayers*. Baltimore, MD: Johns Hopkins University Press.

Johns, S., Kilpatrick, S., Falk, I., & Mulford, B. (2000). Leadership from Within: Rural Community Revitalisation and the School-Community Partnership.

Kane, T. J. (2010). *The price of admission: Rethinking how Americans pay for college*. Washington, DC: Brookings Institution Press.

Kern, A. F. (2009). *Human capital development theory: Implications for education-comparison of influential twenty-first century economists Samuel Bowles and Gary S. Becker*. Retrieved from http://www.personal.psu.edu/afk119/blogs/career_tech_ed/2009/12/human-capital-development-theory.html

Kingston, P., & Lewis, L. (Eds.). (1990). *The high status track: Studies of elite schools and stratification*. Albany, NY: State University of New York Press.

Kohn, M. L., & Schooler, C. (1983). *Work and personality: An inquiry into the impact of social stratification*. New York, NY: Ablex Pub.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American educational research journal*, *32*(3), 465-491.

Leckie, G. (2013). Cross-Classified Multilevel Models - Concepts. LEMMA VLE Module 12, 1-60.

Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 619-678.

Leibenstein, H. (1957). *Economic backwardness and economic growth*. Hoboken, NJ: Wiley,

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of vocational behavior, 45*(1), 79-122.

Leonhardt, M. (2019, Sept). *These 3 states will require you to file a FAFSA to graduate from high school*. CNBC. https://www.cnbc.com/2019/09/30/3-states-will-require-you-to-file-a-fafsa-to-graduate-from-high-school.html

Leung, D. (2006). The male/female earnings gap and female self-employment. *The Journal of Socio-Economics*, *35*(5), 759-779.

Levine, A., and Nidiffer, J. (1996). *Beating the odds: How the poor get to college*. San Francisco, CA: Jossey-Bass.

Light, A., & Strayer, W. (2000). Determinants of college completion: School quality or student ability? *The Journal of Human Resources, 35*(2), 299– 332.

Lin, N. (2001a). Building a network theory of social capital. In N. Lin, K. Cook, and R.S.Burt (eds.), *Social Capital: Theory and Research* (pp. 3–29). New York: Aldine De Gruyter.

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods & Research*, *18*(2-3), 292-326.

Long, M. C. (2008). College quality and early adult outcomes. *Economics of Education review*, *27*(5), 588-602.

Loury, L. D., & Garman, D. (1995). College selectivity and earnings. *Journal of labor Economics*, *13*(2), 289-308.

Lucas, S. R. (2001). Effectively maintained inequality: education transitions, track mobility, and social background effects. *American journal of sociology*, *106*(6), 1642-1690.

Lynch, L. M. (1992). Private-sector training and the earnings of young workers. *The American Economic Review*, *82*(1), 299-312.

MacAllum, K., Glover, D. M., Queen, B., & Riggs, A. (2007). *Deciding on postsecondary education: Final report.* Washington, DC: National Postsecondary Education Cooperative.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science education*, *95*(5), 877-907.

Maple, S. A., & Stage, F. K. (1991). Influences on the choice of math/science major by gender and ethnicity. *American Educational Research Journal*, *28*(1), 37-60.

Marini, M. M., & Fan, P. L. (1997). The gender gap in earnings at career entry. *American Sociological Review*, 588-604.

McCormick, A. C. (2003). Swirling and double-dipping: New patterns of student attendance and their implications for higher education. *New Directions for Higher Education*, *2003*(121), 13-24.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (Vol. 37). CRC press.

McDonough, P. M. (2005). Counseling and college counseling in America's high schools. *State of college admission*, 107-121.

McDonough, P.M. (1997). *Choosing colleges: How social class and schools structure opportunity*. Albany: State University of New York Press.

McDonough, P. M., Antonio, A. L., & Trent, J. W. (1997). Black students, Black colleges: An African American college choice model. *Journal for a Just and Caring Education*, *3*(1), 9-36.

McDonough, P. M., Lising, A., Walpole, A. M., & Perez, L. X. (1998). College rankings: Democratized college knowledge for whom?. *Research in Higher Education*, *39*(5), 513-537.

McDonough, P. M., & Calderone, S. (2006). The meaning of money perceptual differences between college counselors and low-income families about college costs and financial aid. *American Behavioral Scientist*, *49*(12), 1703-1718.

Milburn, A. (2012). *University challenge: How higher education can advance social mobility*. London, United Kindom: Cabinet Office.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of political economy*, *66*(4), 281-302.

Moakler Jr, M. W., & Kim, M. M. (2014). College major choice in STEM: Revisiting confidence and demographic factors. The Career Development Quarterly, 62(2), 128-142.

Monks, J. (2000). The returns to individual and college characteristics: Evidence from the National Longitudinal Survey of Youth. *Economics of Education Review*, *19*(3), 279-289.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.

Mullen, A. L. (2001). Gender, race, and the college science track: Analyzing field concentrations and institutional selectivity. *Journal of Women and Minorities in Science and Engineering*, *7*(4).

Mullen, A. L. (2011). *Degrees of inequality: Culture, class, and gender in American higher education*. JHU Press.

Muskens, M., Frankenhuis, W. E., & Borghans, L. (2019). Low-income students in higher education: Undermatching predicts decreased satisfaction toward the final stage in college. *Journal of youth and adolescence*, *48*(7), 1296-1310.

Musu-Gillette, L., Robinson, J., McFarland, J., KewalRamani, A., Zhang, A., & Wilkinson-Flicker, S. (2016). *Status and Trends in the Education of Racial and Ethnic Groups 2016. NCES 2016-007*. National Center for Education Statistics.

Mykerezi, E., & Mills, B. F. (2008). The wage earnings impact of historically black colleges and universities. *Southern Economic Journal*, 173-187.

National Center for Science and Engineering Statistics (NCSES). (1994). *Women, minorities, and persons with disabilities in science and engineering*. Retrieved from http://www.nsf.gov/statistics/wmpd

National Center for Education Statistics, 2013. *Distribution of earnings and median earnings of persons 25 years old and over, by highest level of educational attainment and sex: 2012.* Retrieved March 2, 2015 from http://nces.ed.gov/programs/digest/d13/tables/dt13_502.40.asp.

National Science Foundation. "Women, Minorities, and Persons with Disabilities in Science and Engineering: May 2008 Updates." 2008b. Retrieved Sept. 22, 2008, from http://www.nsf.gov/statistics/wmpd.

National Science Foundation, Division of Science Resources Statistics. (2007). *S&E degrees, by Race/Ethnicity of recipients: 1995–2004 No NSF 07-308*. Arlington, VA: National Science Foundation

Niu, L. (2017). Family socioeconomic status and choice of stem major in college: An analysis of a national sample. *College Student Journal*, *51*(2), 298-312.

Orfield, G., & Paul, F. G. (1992). State Higher Education Systems and College Completion. Final Report to the Ford Foundation.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. Review of educational research, 66(4), 543-578.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students revisited: A third decade of research* San Francisco, CA: Jossey- Bass.

Paulsen, M. B. (1990). *College choice: understanding student enrollment behavior. ASHE-ERIC Higher Education Report No. 6.* Washington, DC: The George Washington University.

Parent, D. (2000). Industry-specific capital and the wage profile: Evidence from the national longitudinal survey of youth and the panel study of income dynamics. *Journal of Labor Economics*, *18*(2), 306-323.

Perez, T., Cromley, J. G., & Kaplan, A. (2014). The role of identity development, values, and costs in college STEM retention. *Journal of educational psychology*, *106*(1), 315.

Perna, L.W. (2000). Differences in the decision to enroll in college among African Americans, Hispanics, and Whites. *The Journal of Higher Education* 71(2): 117–141.

Perna, L. W. (2005). The key to college access: Rigorous academic preparation. *Preparing for college: Nine elements of effective outreach*, 113-134.

Perna, L. W. (2005). The benefits of higher education: Sex, racial/ethnic, and socioeconomic group differences. *The Review of Higher Education*, *29*(1), 23-52.

Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In *HIGHER EDUCATION:* (pp. 99-157). Springer, Dordrecht.

Perna, L. W., & Titus, M. A. (2004). Understanding differences in the choice of college attended: The role of state public policies. *The Review of Higher Education*, *27*(4), 501-525.

Perna, L. W., Wright-Kim, J., & Jiang, N. (2019). *Questioning the Calculations: Are Colleges Complying with Federal and Ethical Mandates for Providing Students with Estimated Costs*? (Penn AHEAD Research Brief – March 2019). Philadelphia, PA: The Alliance for Higher Education and Democracy (PennAHEAD).

Phillips, M., Yamashiro, K., & Miller, C.E. (2017). *College Readiness Supports In LAUSD High Schools: A First Look*. Retrieved from https://laeri.org/wp-content/uploads/2017/09/laericollegesupports082017.pdf

Polachek, S. W. (1978). Sex differences in college major. *Industrial and Labor Relations Review*, *31*(4), 498-508.

Porter, S. R., & Umbach, P. D. (2006). College major choice: An analysis of person–environment fit. *Research in higher education*, *47*(4), 429-449.

Powell, L. (1990). Factors associated with the underrepresentation of African Americans in mathematics and science. *The Journal of Negro Education*, *59*(3), 292-298.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). *NAEP 2008: Trends in academic progress* (NCES 2009–479). Washington, DC: National Center for Education Statistics.

Rask, K. (2010). Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review*, *29*(6), 892-900.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, 2nd ed. Thousand Oaks, CA: Sage.

Reitz, J. G. (1975). Undergraduate aspirations and career choice: Effects of college selectivity. *Sociology of Education*, 308-323.

Robinson, M. (2003). Student enrollment in high school AP sciences and calculus: How does it correlate with STEM careers?. *Bulletin of Science, Technology & Society*, *23*(4), 265-273.

Roderick, M., Coca, V., & Nagaoka, J. (2011). Potholes on the road to college high school effects in shaping urban students' participation in college application, four-year college enrollment, and college match. *Sociology of Education*, *84*(3), 178-211.

Rodriguez, R. J., & Elbaum, B. (2014). The role of student–teacher ratio in parents' perceptions of schools' engagement efforts. *The Journal of Educational Research*, *107*(1), 69-80.

Roderick, M., Nagaoka, J., Allensworth, E., Coca, V., Correa, M., & Stoker, G. (2006). *From high school to the future: A first look at Chicago public school graduates' college enrollment, college preparation, and graduation from four-year colleges*. Chicago, IL: Consortium on Chicago School Research.

Roderick, M., Nagaoka, J., Coca, V., & Moeller, E. (2008). *From high school to the future: Potholes on the road to college.* Chicago, IL: Consortium on Chicago School Research.

Roderick, M., Nagaoka, J., Coca, V., & Moeller, E. (2009). *From high school to the future: Making hard work pay off. Chicago*, IL: Consortium on Chicago School Research.

Roderick, M., Nagaoka, J., & Coca, V. (2009). College readiness for all: The challenge for urban high schools. *The future of children*, 185-210.

Rosa, M. L. D. L. (2006). Is opportunity knocking? Low-income students' perceptions of college and financial aid. *American Behavioral Scientist*, *49*(12), 1670-1686. Implications for teacher education.

Rosen, S. (1986). The theory of equalizing differences. *Handbook of labor economics*, *1*, 641-692.

Rumberger, R. W., & Thomas, S. L. (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. *Economics of Education Review*, *12*(1), 1-19.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and social psychology bulletin*, *28*(12), 1629-1646.

Sanders, M. G. (2001). The role of" community" in comprehensive school, family, and community partnership programs. *The Elementary School Journal*, *102*(1), 19-34.

Schoeni, R. F. (1995). Marital status and earnings in developed countries. *Journal of population economics*, *8*(4), 351-359.

Schneider, B., and Stevenson, D. (1999). *The ambitious generation: America's teenagers, motivated but directionless.* New Haven, CT: Yale University Press.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American statistical association*, *103*(484), 1334-1344.

Shapiro, C. A., & Sax, L. J. (2011). Major selection and persistence for women in STEM. *New Directions for Institutional Research*, *2011*(152), 5-18.

Shaw, E. J., & Barbuti, S. (2010). Patterns of persistence in intended college major with a focus on STEM majors. *NACADA Journal*, *30*(2), 19-34.

Simpson, J. C. (2001). Segregated by subject: Racial differences in the factors influencing academic major between European Americans, Asian Americans, and African, Hispanic, and Native Americans. *The Journal of Higher Education*, *72*(1), 63-100.

Sjoquist, D. L., & Winters, J. V. (2015). State merit aid programs and college major: A focus on STEM. *Journal of Labor Economics*, *33*(4), 973-1006.

Smart, J. C., Feldman, K. A., & Ethington, C. A. (2006, July). *Holland's theory and patterns of college student success*. Washington, DC: National Postsecondary Education Cooperative.

Smart, J. C., Feldman, K. A., & Ethington, C. A. (2000). *Academic disciplines: Holland's theory and the study of college students and faculty*. Nashville, TN: Vanderbilt University Press.

Smith, J. P. (1984). Race and human capital. *The American Economic Review*, *74*(4), 685-698.

Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review*, *32*, 247-261.

Smith, J., Pender, M., Howell, J., & Hurwitz, M. (2012). *Getting into college: Postsecondary academic undermatch*. New York, NY: The College Board Advocacy & Policy Center.

Smith, T. Y. (2000). 2000 SMET Retention Report: The Retention and Graduation Rates of 1992–98 Entering Science, Mathematics, Engineering and Technology Majors in 119 Colleges and Universities. *University of Oklahoma Center for Institutional Data Exchange and Analysis*.

Spellings, M. (2006). *A test of leadership: Charting the future of US higher education*. US Department of Education.

Stage, F. K., & Hamrick, F. A. (2003). College predisposition at high-minority enrollment, low-income schools. *The Review of Higher Education*, *27*(2), 151-168.

Stampen, J. O., & Fenske, R. H. (1988). The impact of financial aid on ethnic minorities. *The Review of Higher Education*, *11*(4), 337-353.

Staniec, J. F. O. (2004). The effects of race, sex, and expected returns on the choice of college major. *Eastern Economic Journal*, *30*(4), 549-562.

Stewart, Q. T., & Dixon, J. C. (2010). Is it race, immigrant status, or both? An analysis of wage disparities among men in the United States. *International Migration Review*, *44*(1), 173-201.

Stiglitz, J. E. (1976). The efficiency wage hypothesis, surplus labour, and the distribution of income in LDCs. *Oxford economic papers*, *28*(2), 185-207.

St. John, E.P. (2004). Public policy and research on college access: Lessons learned from NCES research and state evaluation studies. In E.P. St. John (ed.), *Public Policy and College Access: Investigating the Federal and State Roles in Equalizing Postsecondary Opportunity* (pp. 181–196). New York: AMS Press, Inc.

Students for fair admissions v. Harvard corp., 397 F. Supp. 3d 126 (D. Mass. 2019).

Titus, M. A. (2006). No college student left behind: The influence of financial aspects of a state's higher education policy on college completion. *The Review of Higher Education*, *29*(3), 293-317.

Thomas, S. L. (2003). Longer-term economic effects of college selectivity and control. *Research in Higher Education*, *44*(3), 263-299.

Thomas, S. L., & Zhang, L. (2005). Post-baccalaureate wage growth within four years of graduation: The effects of college quality and college major. *Research in Higher Education*, *46*(4), 437–459. http://doi.org/10.1007/s11162-005-2969-y

Venezia, A., Kirst, M. W., & Antonio, A. L. (2008). *Betraying the college dream: How disconnected K-12 and postsecondary education systems undermine student aspirations*. Stanford, CA: The Bridge Project.

Watkins, J., & Mazur, E. (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. *Journal of College Science Teaching*, *42*(5), 36-41.

Weeden, K. A. (2002). Why do some occupations pay more than others? Social closure and earnings inequality in the United States. *American Journal of Sociology*, *108*(1), 55-101.

Wellman, B. (1972). Who needs neighbourhoods. *The City: Attacking Modern Myths. McClelland and Stewart, Toronto*, 94-113.

Wiley, C. (1995). Reexamining professional certification in human resource management. *Human Resource Management*, *34*(2), 269-289.

Williams, R. M. (1987). Capital, competition, and discrimination: a reconsideration of racial earnings inequality. *Review of Radical Political Economics*, *19*(2), 1-15.

Winston, G. C., & Hill, C. B. (2005). *Access to the most selective private colleges by high-ability, low-income students: Are they out there?* Williamstown, MA: Williams College.

Witteveen, D., & Attewell, P. (2017). The earnings payoff from attending a selective college. *Social Science Research*, *66*, 154-169.

Wolniak, G. C., Seifert, T. A., Reed, E. J., & Pascarella, E. T. (2008). College majors and social mobility. *Research in Social Stratification and Mobility*, *26*(2), 123-139.

Word, E. R. (1990). *The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report (1985-1990).* Tennessee Board of Education.

Wyner, J. S., Bridgeland, J. M., & DiIulio Jr, J. J. (2007). *Achievementrap: How America is failing millions of high-achieving students from lower-income families.* Los Angeles, CA: Civic Enterprises.

Xu, Y. (2015). Focusing on women in STEM: A longitudinal examination of gender-based earning gap of college graduates. *The Journal of Higher Education*, *86*(4), 489-523.

Zhang, L. (2008). Gender and racial gaps in earnings among recent college graduates. *The Review of Higher Education*, *32*(1), 51-72.

Zhang, L., & Thomas, S. L. (2005). Investments in human capital: Sources of variation in the return to college quality. In *Higher education: Handbook of theory and research* (pp. 241-306). Springer, Dordrecht.